SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

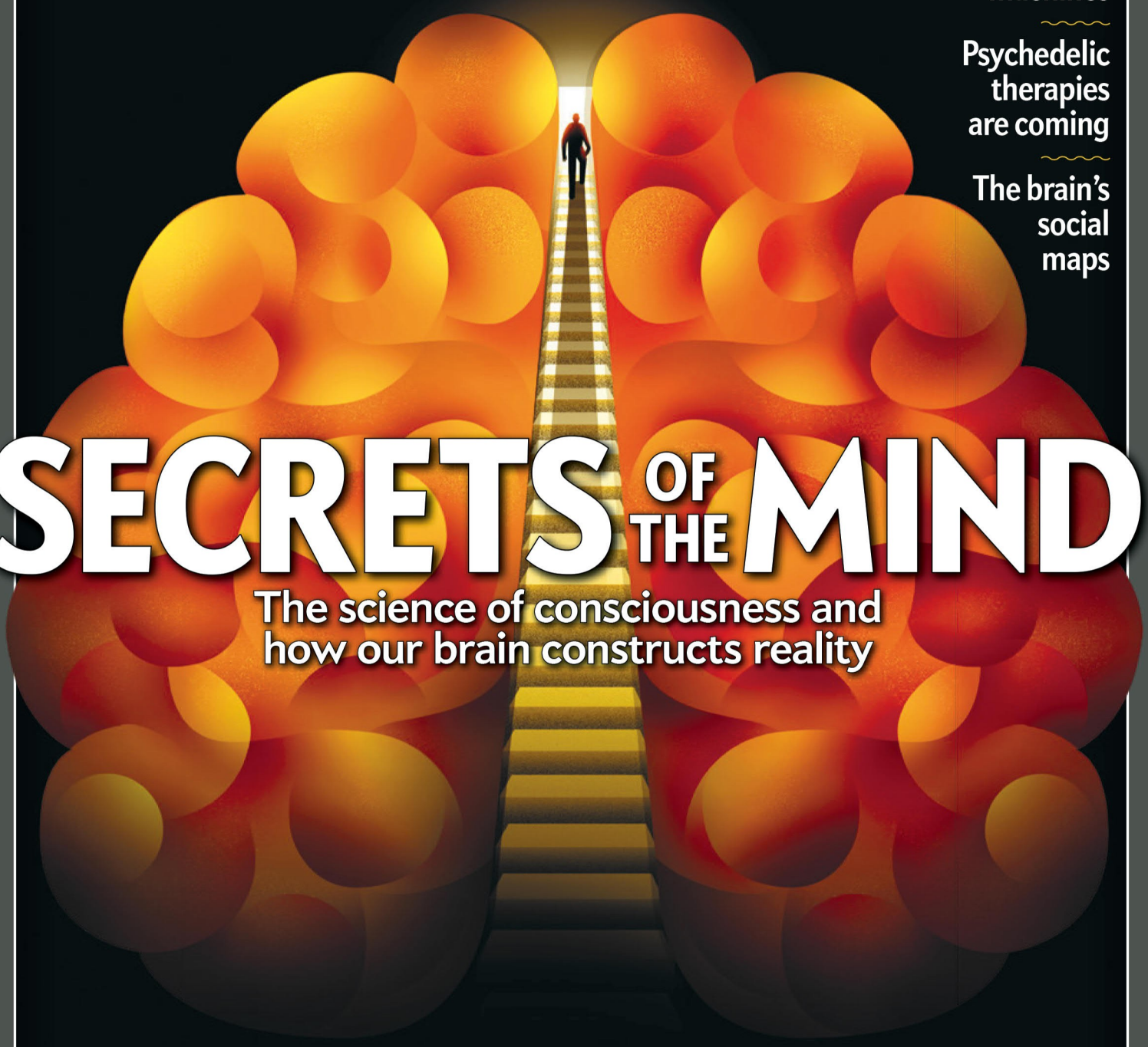
INSIDE

Is free will an illusion?

Mind-reading machines

Psychedelic therapies are coming

The brain's social maps



SECRETS OF THE MIND

The science of consciousness and how our brain constructs reality

WINTER 2022

Scientific American Unlimited

Perfect for science fans, gain access to all of our publications, apps & the full website experience.

Digital archive access back to 1845 including articles by Einstein, Curie and other timeless luminaries in more than 7,000 issues!

12 print and digital issues of *Scientific American* per year

More than 150 eBooks and Collector's Editions

Access to *Scientific American Mind*, *Space & Physics* and *Health & Medicine*

More than 200 articles per month, including news and commentary, on ScientificAmerican.com

sciam.com/unlimited

Scientific American is a registered trademark of Springer Nature America, Inc

SCIENTIFIC AMERICAN®
ESTABLISHED 1845

Secrets of the Mind
is published by the staff
of SCIENTIFIC AMERICAN,
with project management by:

Editor in Chief: **Laura Helmuth**

Managing Editor: **Curtis Brainard**

Senior Editor, Collections:
Andrea Gawrylewski

Creative Director: **Michael Mrak**

Issue Designer: **Lawrence R. Gendron**

Senior Graphics Editor: **Jen Christiansen**

Associate Graphics Editor: **Amanda Montañez**

Photography Editor: **Monica Bradley**

Associate Photo Editor: **Liz Tormes**

Copy Director: **Maria-Christina Keller**

Senior Copy Editors:
Angelique Rondeau, Aaron Shattuck

Managing Production Editor: **Richard Hunt**

Prepress and Quality Manager:
Silvia De Santis

Executive Assistant Supervisor: **Maya Harty**

Acting President: **Stephen Pincock**

Executive Vice President: **Michael Florek**

Vice President, Commercial: **Andrew Douglas**

Publisher and Vice President:
Jeremy A. Abbate

Associate Vice President,
Business Development: **Diane McGarvey**

Marketing Director, Institutional Partnerships
and Customer Development: **Jessica Cole**

Programmatic Product Manager: **Zoya Lysak**

Director, Integrated Media: **Matt Bondlow**

Senior Marketing Manager:
Christopher Monello

Senior Product Manager: **Ian Kelly**

Senior Web Producer: **Jessica Ramirez**

Senior Commercial Operations Coordinator:
Christine Kaelin

Custom Publishing Editor: **Lisa Pallatroni**

Head, Communications, USA: **Rachel Scheer**

Press Manager: **Sarah Hausman**

Production Controller: **Madelyn Keyes-Milch**

Advertising Production Manager:
Michael Broomes

Maria Corte (*cover illustration*);
Jorm Sangsorn/Getty Images (*this page*)

FROM THE EDITOR

Mind from Matter

In 2016 a panel of physicists, a cosmologist and a philosopher gathered at the American Museum of Natural History to discuss an idea seemingly befitting science fiction: Are we living in a computer simulation? How exactly the flesh and blood of our brain is able to formulate an aware, self-examining mind capable of critical thought remains a mystery. Perhaps the answer eludes us because, the panel mused, we are the avatars of a higher species' simulation and simply unable to discover the truth. As intriguing a hypothesis as it is, neuroscience has learned enough about our consciousness to counter such a fantastical possibility.

Newly mapped networks within the human brain show regions that fire in concert to create cognition (*page 6*). Zapping the brain with magnetic pulses while recording neural activity might soon detect conscious thought, which could be especially useful for patients who are awake but unable to communicate or respond to external stimuli (*page 26*). These discoveries chip away at the isolating experience of humanity and the idea that a person can never truly know whether anyone but oneself is truly conscious (*page 32*).

To some extent, we exist in our own bubbles of subjective experience. A growing body of evidence suggests that perception is a construction of the brain (*page 38*). Because the brain initiates some actions before we become aware that we have made a decision, we might even deduce that each of us is some kind of biochemical puppet, but experiments confirm that we do indeed have free will (*page 44*). And our cognition clearly results from highly evolved neural mechanisms, common to all of us, for making new memories (*page 48*), navigating social relationships (*page 54*) and recognizing faces (*page 60*). Ultimately a shared sense of reality influences how we perceive ourselves and the formation of "in-groups" and "out-groups," which can create social and political division (*page 68*).

For all that, however, a lot happens outside of our awareness. Based on neuronal firing patterns, algorithms can infer intended body movements in patients with paralysis of their limbs, illuminating the unconscious brain-body dynamic (*page 74*). During sleep, the brain makes crucial gains in learning, memory and emotion processing (*page 82* and *page 88*).

And there is much out there that can alter our reality, from drugs to disease. Potentially a third of older COVID patients experience delirium, perhaps increasing their risk for dementia later on (*page 116*). Psychiatrists now debate whether bouts of mania in which people report enhanced recall, increased empathy and spirituality might constitute a new category of mental disorder (*page 110*). Ironically, rather than distorting it, substances such as Ecstasy and "magic" mushrooms may actually help restore a rational view of the world (*page 122*).

Consciousness may come from an alien programmer, or perhaps it pervades our universe, not just our brains but all things, as philosopher Philip Goff posits (*page 124*). Either way, our remarkable ability to think about our own thoughts—meta-thoughts, as *Scientific American* columnist John Horgan calls them on page 96—places us at a tantalizing intersection of life and sentience. What we think matters to us regardless of how the mind arises from matter. SA

Andrea Gawrylewski
Senior Editor, Collections, editors@sciam.com

SPECIAL EDITION

SCIENTIFIC AMERICAN

Volume 31, Number 1, Winter 2022

WHAT IS CONSCIOUSNESS?

6 How Matter Becomes Mind
The new discipline of network neuroscience yields a picture of how mental activity arises from carefully orchestrated interactions among different brain areas. *By Max Bertolero and Danielle S. Bassett*

14 How to Make a Consciousness Meter
Zapping the brain with magnetic pulses while measuring its electrical activity is proving to be a reliable way to detect consciousness. *By Christof Koch*

20 Can Lab-Grown Brains Become Conscious?
A handful of experiments are raising questions about whether clumps of cells and disembodied brains could be sentient and how scientists would know if they were. *By Sara Reardon*

26 The Brain Electric
Electrodes that stimulate brain tissue reveal the topography of conscious experience. *By Christof Koch*

32 How Do I Know I'm Not the Only Conscious Being in the Universe?
The solipsism problem, also called the problem of other minds, lurks at the heart of science, philosophy, religion, the arts and the human condition. *By John Horgan*

HOW WE PERCEIVE THE WORLD

38 Our Inner Universes
Reality is constructed by the brain, and no two brains are exactly alike. *By Anil K. Seth*

44 Why We Have Free Will
Neurons fire in your head before you become aware that you have made a decision. But this discovery does not mean you are a "biochemical puppet." *By Eddy Nahmias*

48 The Brain Learns in Unexpected Ways
Neuroscientists have discovered a set of unfamiliar cellular mechanisms for making fresh memories. *By R. Douglas Fields*

54 The Brain's Social Road Maps
Neural circuits that track our whereabouts in space and time may also play vital roles in determining how we relate to other people. *By Matthew Schafer and Daniela Schiller*

60 Face Values
Brain regions that process faces reveal deep insights into the neural mechanisms of vision. *By Doris Y. Tsao*

68 Radical Change
Uncertainty in the world threatens our sense of self. To cope, people embrace populism. *By Michael A. Hogg*

4 WHAT IS CONSCIOUSNESS?

36 HOW WE PERCEIVE THE WORLD

UNCONSCIOUS MIND 72

ALTERED REALITY 100

UNCONSCIOUS MIND

74 The Intention Machine
A new generation of brain-machine interfaces can deduce what a person wants. *By Richard A. Andersen*

82 Infectious Dreams
How the COVID-19 pandemic is changing our sleeping lives. *By Tore Nielsen*

88 Sleep Learning Gets Real
Experimental techniques demonstrate how to strengthen memories when our brains are off-line. *By Ken A. Paller and Delphine Oudiette*

94 Answering Queries in Real Time while Dreaming
Researchers demonstrate that during REM sleep, people can hear—and respond to—simple questions such as "What is eight minus six?" *By Diana Kwon*

96 Can Science Illuminate Our Inner Dark Matter?
Neither introspection nor brain scans can reveal our deepest thoughts. *By John Horgan*

ALTERED REALITY

102 A Disorder of Mind and Brain
A mysterious condition once known as hysteria is challenging the divide between psychiatry and neurology. *By Diana Kwon*

110 The Undiscovered Illness
Hundreds of thousands of people experience mania without ever getting depressed. Why does psychiatry insist on calling them bipolar? *By Simon Makin*

116 Could COVID Delirium Bring On Dementia?
Delirium is very common on COVID wards. Researchers are testing whether these temporary bouts of confusion could bring on permanent cognitive decline. *By Carrie Arnold*

122 A Psychedelic Renaissance
Psilocybin and MDMA represent a first wave of therapies for psychiatric disorders that help patients by changing the way they view reality. *By Danielle Schlosser and Thomas R. Insel*

DEPARTMENTS

1 FROM THE EDITOR
Mind from Matter

124 END NOTE
Does Consciousness Pervade the Universe?
Philosopher Philip Goff answers questions about "panpsychism." *By Gareth Cook*

Articles in this special issue are updated or adapted from previous issues of *Scientific American* and *Nature* and from ScientificAmerican.com.
Copyright © 2022 Scientific American, a division of Springer Nature America, Inc. All rights reserved. Scientific American Special (ISSN 1936-1513), Volume 31, Number 1, Winter 2022, published by Scientific American, a division of Springer Nature America, Inc., 1 New York Plaza, Suite 4600, New York, N.Y. 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, P.O. Box 3187, Harlan Iowa 51537. **Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.**

What Is Consciousness?

How Matter Becomes Mind *page 6*

How to Make a Consciousness Meter *page 14*

Can Lab-Grown Brains Become Conscious? *page 20*

The Brain Electric *page 26*

How Do I Know I'm Not the Only Conscious Being in the Universe? *page 32*

Illustration by Maria Corte

how matter becomes mind

The new discipline of network neuroscience yields a picture of how mental activity arises from carefully orchestrated interactions among different brain areas

By Max Bertolero and Danielle S. Bassett

Illustration by Mark Ross Studios

NETWORKS PERVADE OUR LIVES. EVERY DAY WE USE INTRICATE NETWORKS of roads, railways, maritime routes and skyways traversed by commercial flights. They exist even beyond our immediate experience. Think of the World Wide Web, the power grid and the universe, of which the Milky Way is an infinitesimal node in a seemingly boundless network of galaxies. Few such systems of interacting connections, however, match the complexity of the one underneath our skull.

Neuroscience has gained a higher profile in recent years, as many people have grown familiar with splashily colored images that show brain regions "lighting up" during a mental task. There is, for instance, the temporal lobe, the area by your ear, which is involved with memory, and the occipital lobe at the back of your head, which dedicates itself to vision.

What has been missing from this account of human brain function is how all these distinct regions interact to give rise to who we are. Our laboratory and others have borrowed a language from a branch of mathematics called graph theory that allows us to parse, probe and predict complex interactions of the brain that bridge the seemingly vast gap between frenzied neural electrical activity and an array of cognitive tasks—sensing, remembering, making decisions, learning a new skill and initiating movement. This new field of network neuroscience builds on and reinforces the idea that certain regions of the brain carry out defined activities. In the most fundamental sense, what the brain is—and thus who we are as conscious beings—is, in fact, defined by a sprawling network of 100 billion neurons with at least 100 trillion connecting points, or synapses.

Network neuroscience seeks to capture this complexity. We can now model the data supplied by brain imaging as a graph composed of nodes and edges. In a graph, nodes represent the units of the network, such as neurons or, in another context, airports. Edges serve as the connections between nodes—think of one neuron intertwined with the next or contemplate airline flight routes. In our work, the human brain is reduced to a graph of roughly 300 nodes. Diverse areas can be linked together by edges representing the brain's structural connections: thick bundles of tubular wires called white matter tracts that tie together brain regions. This depiction of the brain as a unified network has already furnished a clearer picture of cognitive functioning, along with the practical benefit of enabling better diagnoses and treatment of psychiatric disorders. As we glimpse ahead, an understanding of brain networks may lead to a blueprint for improved artificial intelligence, new medicines and electrical-stimulation technology to alter malfunctioning neural circuitry in depression—and perhaps even the development of genetic therapies to treat mental illness.

THE MUSIC OF THE MIND

TO UNDERSTAND HOW networks underlie our cognitive capabilities, first consider the analogy of an orchestra playing a symphony. Until recently, neuroscientists have largely studied the functioning of individual brain regions in isolation, the neural equivalent of separate brass, percussion, string and woodwind sections. In the brain, this stratification represents an approach that dates back to Plato—quite simply, it entails carving nature at the joints and then studying the individual components that remain.

Just as it is useful to understand how the amygdala helps to process emotions, it is similarly vital to grasp how a violin produces high-pitched sounds. Still, even a complete list of brain regions and their functions—vision, motor, emotion, and so on—does not tell us how the brain really works. Nor does an inventory of instruments provide a recipe for Beethoven's *Eroica* symphony.

Network neuroscientists have begun to tame these mysteries by examining the way each brain region is embedded in a larger network of such regions and by mapping the connections between regions to study how each is embedded in the large, integrated network that is the brain. There are two major approaches. First, examining structural connectivity captures the instrumentation of the brain's orchestra. It is the physical means of creating the music, and the unique instrumentation of a given musical work constrains what can be played. Instrumentation matters, but it is not the music itself. Put another way, just as a collection of instruments is not music, an assemblage of wires does not represent brain function.

Second, living brains are massive orchestras of neurons that fire together in quite specific patterns. We hear a brain's music by measuring the correlation between the activity of each pair of regions, indicating that they are working in concert. This measure of joint activity is known as functional connectivity, and we colloquially think of it as reflecting the music of the brain. If two regions fire with the same time-varying fluctuations, they are considered to be functionally connected. This music is just as important as the decibels produced by a French horn or viola. The volume of the brain's music can be thought of as the level of activity of electrical signals buzzing about one brain area or another.

At any moment, though, some areas within the three-pound

organ are more active than others. We have all heard the saying that people use a small fraction of their brain capacity. In fact, the entire brain is active at any point in time, but a given task modulates the activity of only a portion of the brain from its baseline level of activity.

That arrangement does not mean that you fulfill only half of your cognitive potential. In fact, if your entire brain were strongly active at the same time, it would be as if all the orchestra members were playing as loudly as possible—and that scenario would create chaos, not enable communication. The deafening sound would not convey the emotional overtones present in a great musical piece. It is the pitch, rhythms, tempo and strategic pauses that communicate information, both during a symphony and inside your head.

MODULARITY

JUST AS AN ORCHESTRA can be divided into groups of instruments from different families, the brain can be separated into collections of nodes called modules—a description of localized networks. All brains are modular. Even the 302-neuron network of the nematode *Caenorhabditis elegans* has a modular structure. Nodes within a module share stronger connections to one another than to nodes in other modules.

Each module in the brain has a certain function, just as every family of instruments plays a role in the symphony. We recently performed an evaluation of a large number of independent studies—a meta-analysis—that included more than 10,000 functional magnetic resonance imaging (fMRI) experiments of subjects performing 83 different cognitive tasks and discovered that separate tasks map to different brain-network modules. There are modules occupied with attention, memory and introspective thought. Other modules, we found, are dedicated to hearing, motor movement and vision.

These sensory and motor cognitive processes involve single, contiguous modules, most of which are confined to one lobe of the brain. We also found that computations in modules do not spur more activity in other modules—a critical aspect of modular processing. Imagine a scenario in which every musician in an orchestra had to change the notes played every time another musician changed his or her notes. The orchestra would spiral out of control and would certainly not produce aesthetically pleasing sounds. Processing in the brain is similar—each module must be able to function mostly independently. Philosophers as early as Plato and as recent as Jerry Fodor have noted this necessity, and our research confirms it.

Even though brain modules are largely independent, a symphony requires that families of instruments be played in unison. Information generated by one module must eventually be integrated with other modules. Watching a movie with only a brain module for vision—without access to the one for emotions—would detract greatly from the experience.

For that reason, to complete many cognitive tasks, modules must often work together. A short-term memory task—holding a new phone number in your head—requires the cooperation of auditory, attention and memory-processing modules. To integrate and control the activity of multiple modules, the brain uses hubs—nodes where connections from the brain's different modules meet.

Some key modules that control and integrate brain activity are less circumspect than others in their doings. Their connections extend globally to multiple brain lobes. The frontoparietal control module spans the frontal, parietal and temporal lobes. It developed relatively recently on the timescale of evolution. The module is especially large in humans, relative to our closest primate ancestors. It is analogous to an orchestra conductor and becomes active across a large number of cognitive tasks.

The frontoparietal module ensures that the brain's multiple modules function in unison. It is heavily involved in what is called executive function, which encompasses the separate processes of decision-making, short-term memory and cognitive control. The last is the ability to develop complex strategies and inhibit inappropriate behavior.

Another highly interconnected module is the salience module, which hooks up to the frontoparietal control module and contributes to a range of behaviors related to attention and responding to novel stimuli. For example, take a look at two words: blue and red. If you are asked to respond with the color of the word, you will react much faster to the one set in red. The frontoparietal and salience modules activate when responding to the color green because you have to suppress a natural inclination to read the word as "blue."

Finally, the default mode module spans the same lobes as the frontoparietal control network. It contains many hubs and is linked to a variety of cognitive tasks, including introspective thought, learning, memory retrieval, emotional processing, inference of the mental state of others and even gambling. Critically, damage to these hub-rich modules disturbs functional connections throughout the brain and causes widespread cognitive difficulties, just as bad weather at a hub airport delays air traffic all over the country.

PERSONAL CONNECTIONS

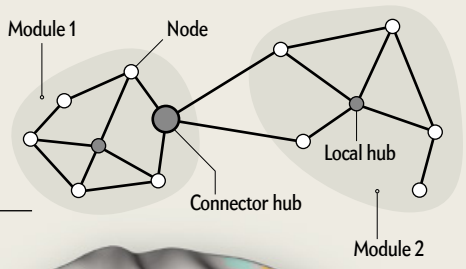
ALTHOUGH OUR BRAINS have certain basic network components—modules interconnected by hubs—each of us shows slight variations in the way our neural circuits are wired. Researchers have devoted intense scrutiny to this diversity. In an initial phase of what is called the Human Connectome Project, 1,200 young people volunteered to participate in a study of brain-network architecture, funded by the National Institutes of Health. (The final goal of the project is to cover the entire life span.) Each individual's structural and functional connectivity networks were probed using fMRI. These data were supplemented by a cognitive battery of testing and questionnaires to analyze 280 behavioral and cognitive traits. Participants provided information about how well they slept, how often they drank alcohol, their language and memory abilities, and their emotional states. Neuroscientists from all over the world have been poring over this incredibly rich data set to learn how our brain networks encode who we are.

Using data from hundreds of participants in the Human Connectome Project, our lab and others have demonstrated that brain-connectivity patterns establish a "fingerprint" that distinguishes each individual. People with strong functional connections among certain regions have an extensive vocabulary and exhibit higher fluid intelligence—helpful for solving novel problems—and are able to delay gratification. They tend to have more education and life satisfaction and better memory and attention. Others with weaker functional connections among those same brain areas have lower fluid intelligence, histories of substance use, poor sleep and a decreased capacity for concentration.

Inspired by this research, we showed that the findings could

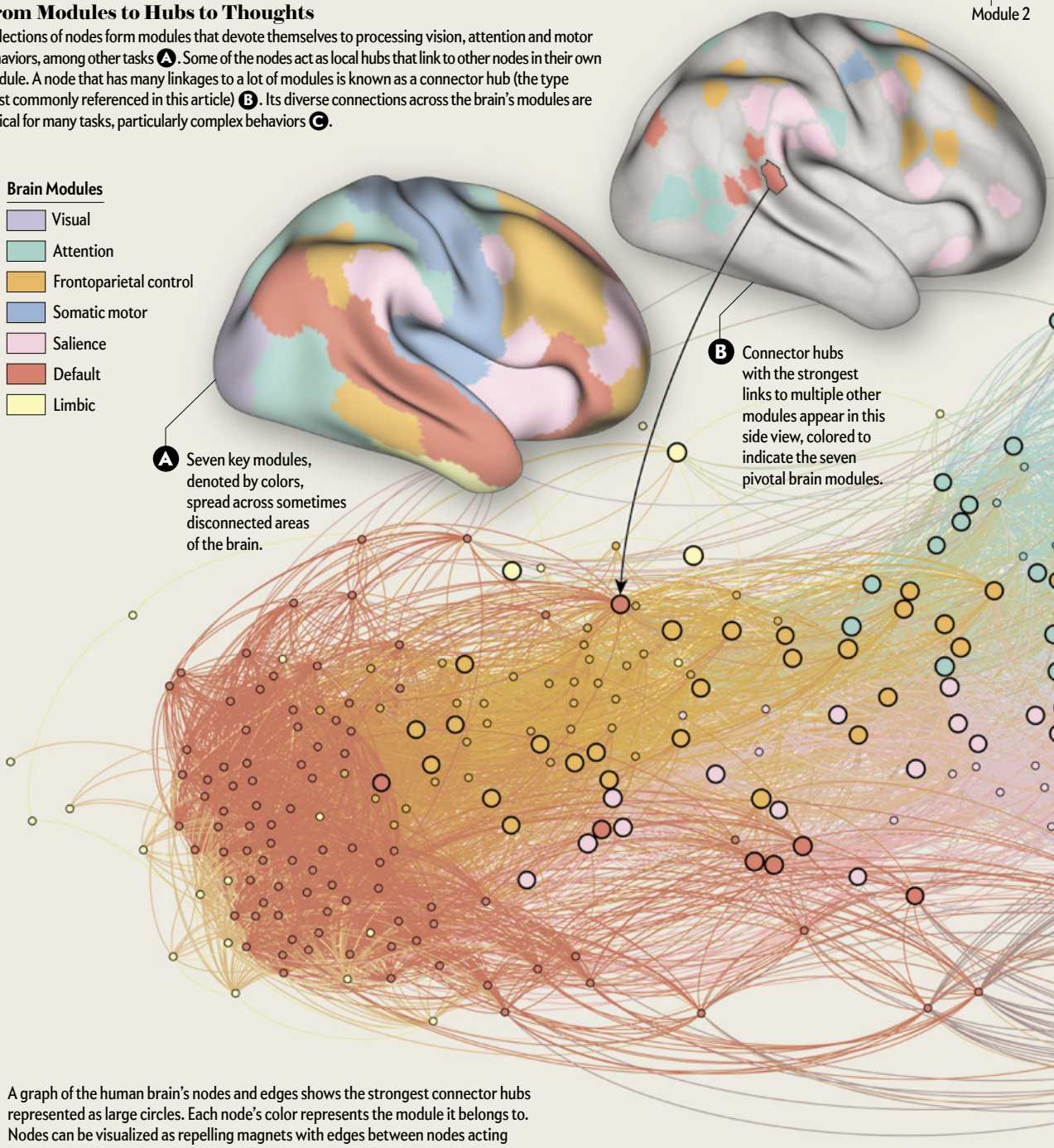
Decoding 100 Trillion Messages

The Milky Way has hundreds of billions of stars—just a fraction of the 100 trillion connections in our brains that enable us to sense, think and act. To unravel this complexity, network neuroscientists create a map, or "graph," consisting of nodes linked by edges that fit into modules, which are tethered to one another with highly connected nodes called hubs.



From Modules to Hubs to Thoughts

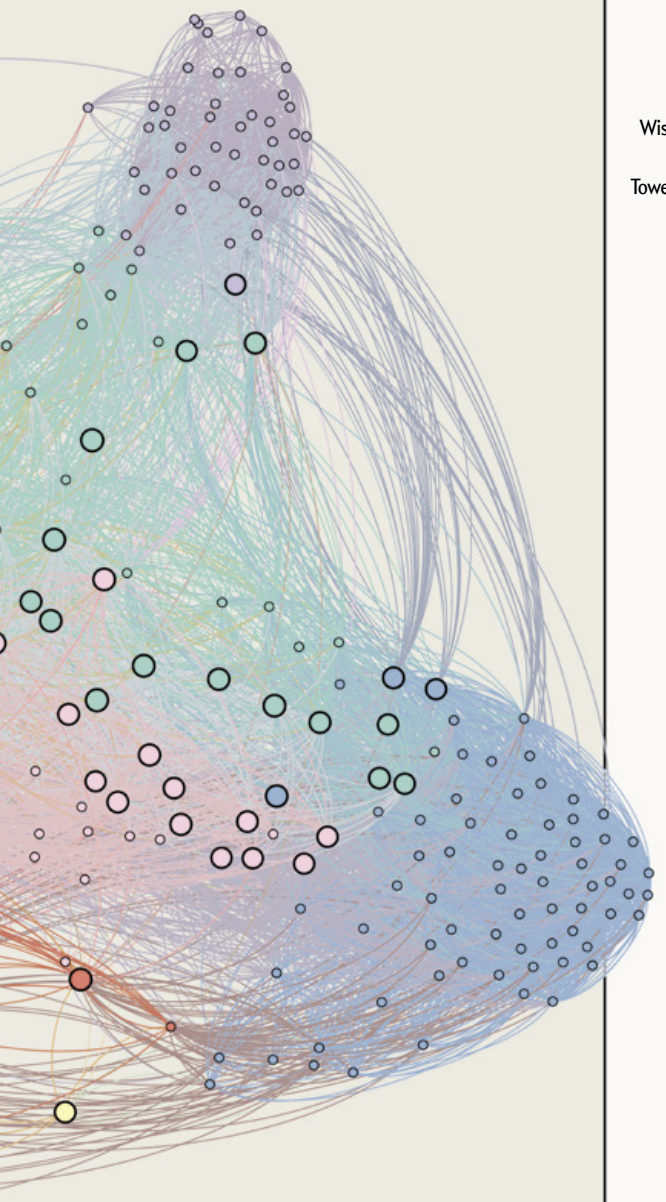
Collections of nodes form modules that devote themselves to processing vision, attention and motor behaviors, among other tasks **A**. Some of the nodes act as local hubs that link to other nodes in their own module. A node that has many linkages to a lot of modules is known as a connector hub (the type most commonly referenced in this article) **B**. Its diverse connections across the brain's modules are critical for many tasks, particularly complex behaviors **C**.

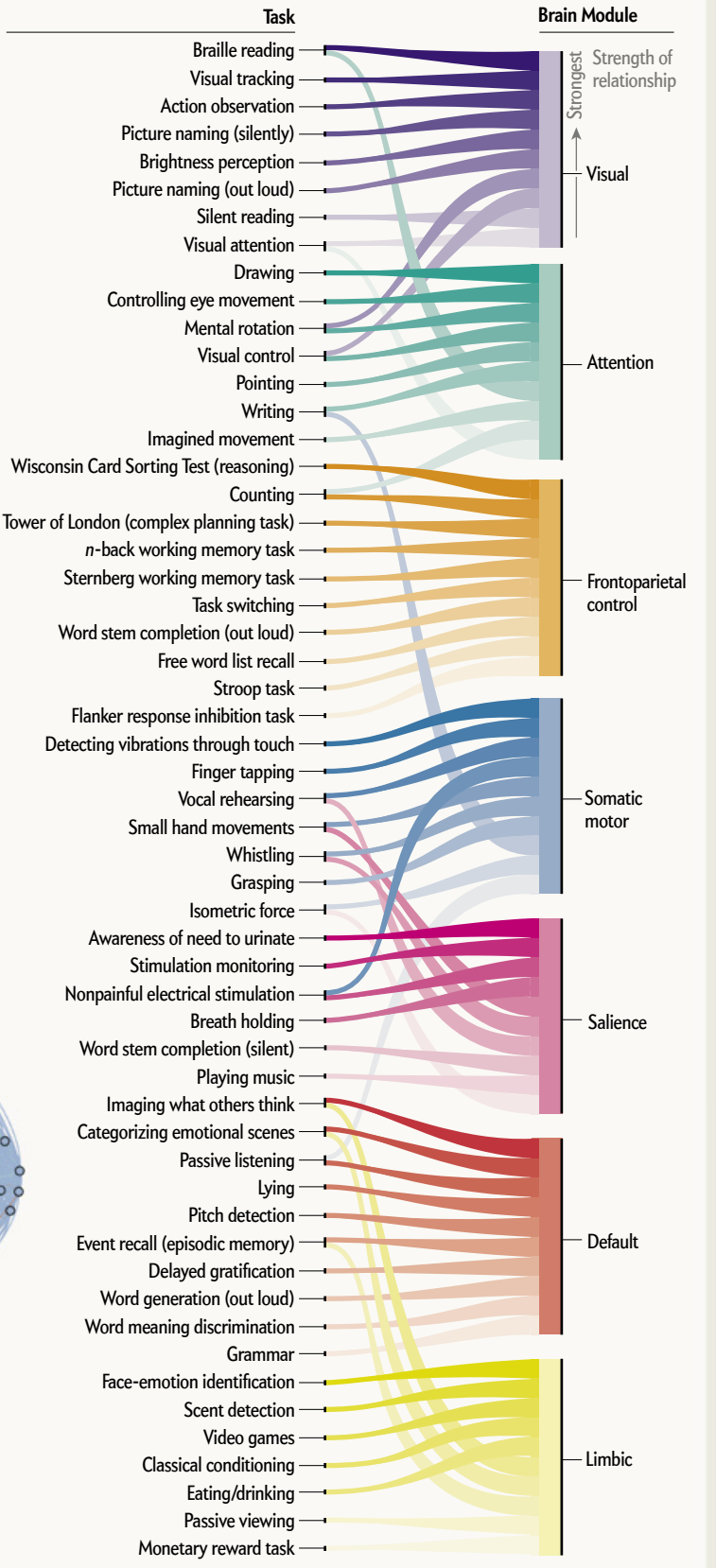

Brain Modules
Visual
Attention
Frontoparietal control
Somatic motor
Salience
Default
Limbic
A Seven key modules, denoted by colors, spread across sometimes disconnected areas of the brain.
B Connector hubs with the strongest links to multiple other modules appear in this side view, colored to indicate the seven pivotal brain modules.

C A graph of the human brain's nodes and edges shows the strongest connector hubs represented as large circles. Each node's color represents the module it belongs to. Nodes can be visualized as repelling magnets with edges between nodes acting as springs that hold them together. Tightly connected nodes cluster together. Connector hubs occupy the center because they are well connected to all modules.

Putting It Together

Modules for vision, attention and other cognitive functions are dedicated to specific tasks, often represented here by psychological tests. The most active tasks rise to the top. The visual module, for instance, is involved with naming, reading and observing. Many tasks require multiple modules. For example, a mental rotation task recruits both the visual and the attention modules. Some modules are entrusted with more abstract tasks. The frontoparietal module engages in switching tasks or recalling lists. The default mode module attends to subjective emotional states or passive listening when a person is at rest.

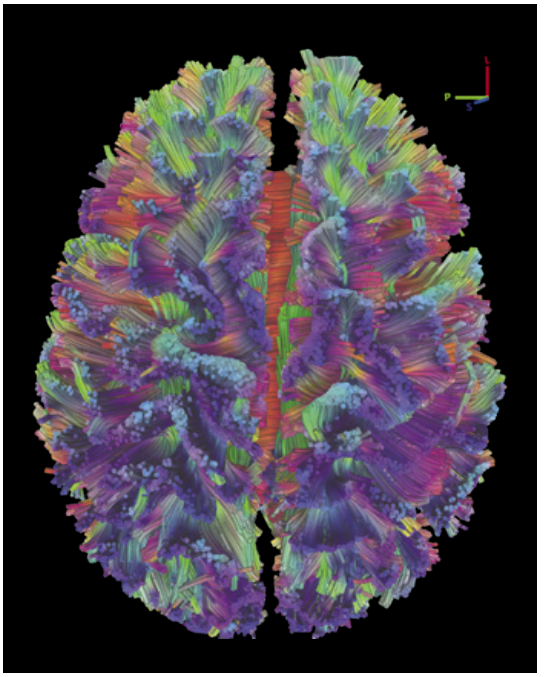




Graphics by Max Bertolero (brains and network diagram) and Jen Christiansen (task chart)

be described by particular patterns among the hub connections. If your brain network has strong hubs with many connections across modules, it tends to have modules that are clearly segregated from one another, and you will perform better on a range of tasks, from short-term memory to mathematics, language or social cognition. Put simply, your thoughts, feelings, quirks, flaws and mental strengths are all encoded by the specific organization of the brain as a unified, integrated network. In sum, it is the music your brain plays that makes you *you*.

The brain's synchronized modules both establish your identity and help to retain it over time. The musical compositions they play appear to always be similar. The likeness could be witnessed when participants in two other studies in the Human Connectome Project engaged in various tasks that involved short-term memory, recognition of the emotions of others, gambling, finger tapping, language, mathematics, social reasoning and a self-induced "resting state" in which they let their mind wander.



MULTITUDES of white matter connections in this scan are used to model the brain's physical pathways—functional networks use these structural linkages to carry out an array of cognitive tasks.

Fascinatingly, the networks' functional wiring has more similarities than expected across all these activities. Returning to our analogy, it is not as if the brain plays Beethoven when doing math and Tupac when resting. The symphony in our head is the same musician playing the same musical genre. This consistency derives from the fact that the brain's physical pathways, or structural connections, place constraints on the routes over the brain's integrated network that a neural signal can travel. And those pathways delineate how functional connections—the ones, say, for math or language—can be configured. In the musical metaphor, a bass drum cannot play the melodic line of a piano.

Changes in the brain's music inevitably occur, just as new arrangements do for orchestral music. Physical connections undergo alterations over the course of months or years, whereas functional connectivity shifts on the order of seconds, when a person switches between one mental task and the next.

Transformations in both structural and functional connectivity are important during adolescent brain development, when the finishing touches of the brain's wiring diagram are being refined. This period is of critical importance because the first signs of mental disorders often appear in adolescence or early adulthood.

One area our research relates to is understanding how brain networks develop through childhood and adolescence and into adulthood. These processes are driven by underlying physiological changes, but they are also influenced by learning, exposure to new ideas and skills, an individual's socioeconomic status and other experiences.

Brain-network modules emerge very early in life, even in the womb, but their connectivity is refined as we grow up. Consistent strengthening of the structural connections to hubs throughout the course of childhood is associated with an increase in the segregation between modules and an augmentation in the efficiency with which young people perform executive tasks such as complex reasoning and self-regulation. We have also found that the segregation of modules from one another is more rapid in children who have a higher socioeconomic status, highlighting the key impact of their environment.

Although changes in structural connectivity are slow, the reconfiguration of functional connections can occur quickly, in a few seconds or minutes. These rapid shifts are instrumental for moving between tasks and for the massive amount of learning demanded even by a single task. In a set of studies that we published from 2011 to 2019, we found that networks with modules that can change readily turn up in individuals who have greater executive function and learning capacity.

To better understand what was happening, we used publicly available data from a landmark study known as MyConnectome, in which Stanford University psychology professor Russell Poldrack personally underwent imaging and cognitive appraisals three times a week for more than a year. Whereas modules are mostly autonomous and segregated, at times the brain will spontaneously reorganize its connections. This property, called functional network flexibility, lets a node with strong functional connections within a module suddenly establish many connections to a different module, changing the flow of information through the network. Using data from this study, we found that the rerouting of a network's connections changes from day to day in a manner that matches positive mood, arousal and fatigue. In healthy individuals, such network flexibility correlates with better cognitive function.

DISSONANT NOTES

THE CONFIGURATION of brain connections also reflects one's mental health. Aberrant connectivity patterns accompany depression, schizophrenia, Alzheimer's, Parkinson's, autism spectrum disorder, attention deficit disorder, dementia and epilepsy.

Most mental illnesses are not confined to one area of the brain. The circuitry affected in schizophrenia extends quite widely across the entire organ. The so-called disconnectivity hypothesis for schizophrenia holds that there is nothing abnormal about the individual modules. Instead the disarray relates to an overabundance of connections between modules.

In a healthy brain, modules are mostly autonomous and segregated, and the ability to bring about flexible changes in network connections is beneficial for cognitive functioning—with-

Danielle S. Bassett and Matthew Cieslak/University of Pennsylvania

in certain limits. In our lab, we found that in the brains of people with schizophrenia and their first-degree relatives, there is an overabundance of flexibility in how networks reconfigure themselves. Auditory hallucinations might result when nodes unexpectedly switch links between speech and auditory modules. The uninvited mix can result in what seem to be the utterings of voices in one's head.

Like schizophrenia, major depressive disorder is not caused by a single abnormal brain region. Three specific modules appear to be affected in depression: the frontoparietal control, salience and default mode modules. In fact, the symptoms of depression—emotional disinhibition, altered sensitivity to emotional events and rumination—map to these modules.

As a result, normal communication among the three modules becomes destabilized. Activities from module to module typically tug back and forth to balance the cognitive processing of sensory inputs with more introspective thoughts. In depression, though, the default mode dominates, and the afflicted person lapses into ruminative thought. The music of the brain thus becomes increasingly unbalanced, with one family of instruments governing the symphony. These observations have broadened our understanding of the network properties of depression to the extent that a connectivity pattern in a brain can allow us to diagnose certain subtypes of the disorder and determine which areas should be treated with electrical-stimulation technology.

NETWORKS EVOLVE

BESIDES STUDYING DEVELOPMENT, network neuroscientists have begun to ask why brain networks have taken their present form over tens of thousands of years. The areas identified as hubs are also the locations in the human brain that have expanded the most during evolution, making them up to 30 times the size they are in macaques. Larger brain hubs most likely permit greater integration of processing across modules and so support more complex computations. It is as if evolution increased the number of musicians in a section of the orchestra, fostering more intricate melodies.

Another way neuroscientists have explored these questions is by creating computer-generated networks and subjecting them to evolutionary pressures. In our lab, we have probed the evolutionary origins of hubs. This exercise started with a network in which all edges were placed uniformly at random. Next, the network was rewired, mimicking natural selection to form segregated modules and display a property known in network science as small-worldness, in which paths form to let distant network nodes communicate with surprising ease. Thousands of such networks then evolved, each of which ultimately contained hubs strongly connected to multiple modules but also tightly interconnected to one another, forming what is called a club. Nothing in the selection process explicitly selected for a club of hubs—they simply emerged from this iterative process.

This simulation demonstrates that one potential solution to evolving a brain capable of exchanging information among modules requires hubs with strong connections. Notably, real networks—brains, airports, power grids—also have durable, tightly interconnected hubs, exactly as predicted by evolutionary experiments. That observation does not mean evolution necessarily occurred in the same way as the simulation, but it shows a possible means by which one of nature's tricks might operate.

STATES OF MIND

WHEN NOBEL PRIZE–WINNING PHYSICIST Richard Feynman died in 1988, his blackboard read, "What I cannot create, I do not understand." He created a beautiful aphorism, yet it misses a pivotal idea: it should be revised to "What I cannot create *and control,* I do not understand." Absent such control, we still know enough to enjoy a symphony, even if we do not qualify to be the conductor.

When it comes to the brain, we have a basic understanding of its form and the importance of its network architecture. We know that our brain determines who we are, but we are just beginning to understand how it all happens. To rephrase mathematician Pierre-Simon Laplace's explanation of determinism and mechanics and apply it to the brain, one's present brain, and so one's mental state, can be thought of as a compilation of past states that can be used to predict the future. A neuroscientist who knew all the principles of brain function and everything about someone's brain could predict that person's mental conditions—the future, as well as the past, would be present inside the person's mind.

This knowledge could be used to prevent pain and suffering, given that many mental illnesses are associated with network abnormalities. With enough engineering ingenuity, we may develop implanted devices that alter or even generate new brain networks or edit genomes to prevent the disorganized networks associated with mental disorders from occurring in the first place. Such an achievement would enable us to treat diseases and to restore brain function after stroke or injury and enhance it in healthy individuals.

Before those futuristic scenarios materialize, two major gaps must be filled: we need to know more about how personal genetics, early-life development and environment determine one's brain's structure and how that structure leads to functional capacities. Neuroscientists have some knowledge from the human genome about the structure that gives rise to functional networks but still need to learn precisely how this process occurs. We are starting to grasp the way brain networks develop and are shaped by the environment but are not close to explaining the entire complexity of this process. The brain's wiring, its structural connectivity, constrains how various modules interact with one another, but our knowledge remains limited. As we fill in these gaps, chances improve for interventions to guide brain functioning into healthy trajectories.

What holds us back, for the moment, is our still blurry vision of the brain—it is as if we are outside the concert hall and have seen only sketches of the instruments. Inside each brain region that neuroscientists study are millions of neurons firing every millisecond, and we are able just to indirectly measure their average activity levels every second or so. Thus far we can roughly identify the human brain's structural connections. Luckily, scientists and engineers have taken steps to deliver ever clearer data that will enable a deeper look into perhaps the most complex network in the known universe: your brain. SA

Max Bertolero is a research scientist at the University of Pennsylvania. He received a doctorate in systems neuroscience from the University of California, Berkeley, and undergraduate degrees in philosophy and psychology from Columbia University.

Danielle S. Bassett is an associate professor in the department of bioengineering at the University of Pennsylvania, where she studies networks in physical and biological systems. In 2014 she became a MacArthur Fellow.

HOW TO MAKE A CONSCIOUSNESS METER

Zapping the brain with magnetic pulses while measuring its electrical activity is proving to be a reliable way to detect consciousness

By Christof Koch

Illustration by Ashley Mackenzie

HAVE DIED MANY TIMES OVER. EVERY NIGHT WHEN I LAY DOWN MY WEARY SELF to rest, my consciousness is extinguished. I experience nothing until I wake up inside my sleeping body—in a dream disconnected from the external world. Or later consciousness resurfaces in the morning on my return to the wakening world.

Daily life contains many such experiences. In my childhood, I had an appendectomy and was anesthetized—my consciousness was switched off and, following the surgery, was restored. A fading memory from my teenage years places me in the back seat of a Renault that is driving down a tree-lined avenue in North Africa. Suddenly, the scenery changes abruptly. I'm on the same street, seeing things now from the ground up. The car had hit a tree, ejecting me onto the cobblestones, and I lost consciousness.

Many readers will have had similar recollections of consciousness lost and regained. We are used to the diurnal cycle of waking, sleeping and dreaming. But that experience is not the same for everyone. For some patients with brain trauma, consciousness flees for days, weeks or longer.

In practice, a clinician may have difficulty establishing whether someone is quietly sleeping, anesthetized or severely brain-injured. Is a person lying with eyes open experiencing anything, no matter the content, or has the conscious mind fled the body and left no one at home?

Ideally, a technology could be devised to serve as a form of consciousness meter to answer these questions. At first, the idea of the equivalent of a blood pressure cuff for consciousness might seem absurd. But the development of several new technologies has raised real prospects for detectors that meet the criteria for consciousness meters—devices useful in medical or research settings to determine whether a person is experiencing anything at all. This ability to detect consciousness could also help physicians and family members make critical decisions, such as withdrawal of life-sustaining therapy, for tens of thousands of uncommunicative patients.

RECORDING BRAIN WAVES

CONTEMPLATING THE possibility of a consciousness meter requires consideration of the internal dynamics of our mental life, activity that waxes and wanes within fractions of a second, dictating the measuring of those fluctuating brain signals at a similar timescale. The most important physiological tool to infer consciousness from probing the brain has been, and continues to be, the electroencephalogram (EEG).

The EEG was developed by German psychiatrist Hans Berger, whose lifelong quest was to uncover the link between objective brain activity and subjective phenomena. He recorded the first ever brain waves of a patient in 1924 but, filled with doubt, did not publish his findings until 1929. The rest is history, as the EEG became the foundational tool of an entire field of medicine called clinical neurophysiology, although Berger was never accorded any significant recognition in Nazi Germany and hanged himself in 1941, despite being nominated for the Nobel Prize several times.

There are, of course, other ways to record brain activity besides the venerable EEG. The most common tools measure the dynamics of blood flowing inside the brain with magnetic scanners or track the magnetic field around the brain with magnetoencephalography (MEG). Yet these instruments, as well as more recent techniques such as near-infrared spectroscopy, come with methodological and practical issues that preclude their routine clinical use for the time being.

The EEG measures the tiny voltage fluctuations (10 to 100 microvolts) generated by electrical activity across the neocortex, the brain's outer surface, which is responsible for perception, action, memory and thought. The main actors whose collective electrical activity is thought to be responsible for the EEG signals, via a mechanism known as volume conduction, are cortical pyramidal neurons, named for their tetrahedral shape. Contributions from deeper structures, such as the thalamus, have to be inferred indirectly via their action on cortical cells. The technology relies on electrodes placed directly on the scalp—that is, without the need for invasive surgery to penetrate the skull. With the move toward high-density EEG setups—with up to 256 electrodes—maps showing the distribution of electrical activity across the brain have become commonplace.

Still, placing the electrodes with their wet, conductive gel onto the scrubbed skin of the head is cumbersome, time-

consuming, and prone to errors if the electrodes move, all of which limits the technology. With today's more sensitive dry electrodes, the EEG is morphing from a clinical tool into a consumer device that can be used for biofeedback—allowing athletes or do-it-yourself "brain hackers" to focus their thoughts or insomniacs to track, deepen and extend their naturally occurring sleep.

From the late 1940s onward, detection of an "activated EEG" signal was the surest sign of a conscious subject. This state is characterized by low-voltage, rapid up-and-down fluctuating waves that are desynchronized rather than in lockstep across the skull. In general, as the EEG shifts to lower frequencies, consciousness is less likely to be present. Yet there are enough exceptions to this rule that it cannot serve as a general basis to diagnose absence or presence of consciousness in a given individual. Thus, scientists and clinicians alike have cast about for more reliable measures and have now found one based on a fundamental property of any conscious experience.

INTO THE NETHERWORLD

BEFORE WE COME to that, we should consider why clinicians care about detecting consciousness in two distinct groups of patients (pediatric patients represent a different challenge that will not be addressed here). The first consists of adults with severe disorders of consciousness following traumatic brain injury caused by gunshots, falls, accidents, and so on, infections of the brain (encephalitis) or its surrounding protective layers (meningitis), stroke, or drug or alcohol intoxication. After surviving the initial insult, patients are stable but disabled and bedridden, unable to speak or signal their thoughts and intentions. With proper nursing care to avoid bedsores and infections, these patients can live for years.

In this first group, clinicians distinguish several subcategories. Patients in a vegetative state, which is better described by the less pejora-

Zapping and Zipping

In pursuit of a consciousness test, Silvia Casarotto of the University of Milan and her colleagues recruited 102 nonbrain-injured subjects and 48 still responsive and awake brain-injured patients. Their brains were "zapped" with magnetic pulses (transcranial magnetic stimulation) in both conscious and unconscious states, and brain activity was detected with an EEG and analyzed with a data-compression algorithm—and so it was said to be "zipped." A value known as a perturbational complexity index (PCI) was calculated for the EEGs—and participants were also interviewed about their state of mind. It was determined that a conscious person exhibited at least one value above 0.31 (PCI*), whereas unconscious subjects all had lower scores. Using this value, the zap-and-zip testing was then performed on patients with severe disorders of consciousness (*results not shown*), finding some individuals who appeared to be conscious.

Source: "Stratification of Unresponsive Patients by an Independently Validated Index of Brain Complexity," by Silvia Casarotto et al., in *Annals of Neurology*, Vol. 80, No. 5; November 2016

Graphic by Amanda Montañez

tive term "unresponsive wakefulness syndrome" (UWS), cycle in and out of sleep. Yet setting up a bedside communication channel—"if you hear me, squeeze my hand or move your eyes"—meets with failure. UWS patients do swallow, yawn, and open and move their eyes or head but not in a seemingly intentional manner. No willed actions are left—only brain stem reflexes, activity that controls basic processes such as breathing, sleep-wake transitions, heart rate, eye movements and pupillary responses. Terri Schiavo is a name many remember, a patient in Florida who, following cardiac arrest, was resuscitated and lingered for 15 years in UWS until her medically induced death in 2005. UWS patients are a modern phenomenon depending for their survival on the infrastructure of 911, emergency helicopters and advanced medical care. There are more than 10,000 such individuals in the U.S. alone, living in hospices or nursing homes or at home.

Whereas behavioral evidence is compatible with the notion that UWS patients do not experience anything, it is important to recall that "absence of evidence is not evidence of absence" and to give the patients the benefit of doubt. There is a diagnostic gray zone into which UWS patients fall as to the question of whether their injured brains are capable of experiencing pain, distress, anxiety, isolation, quiet resignation, a full-blown stream of thought—or perhaps just nothing. Some studies have suggested that 20 percent of UWS patients are conscious and are therefore misdiagnosed. To family and friends who may care for their loved one for years, knowing whether anybody is mentally there can make a dramatic difference.

The situation is less ambiguous for minimally conscious state (MCS) patients. Unable to speak, they can signal but often only in a sparse, minimal and erratic fashion, smiling or crying in appropriate emotional situations, vocalizing or gesturing on occasion, or tracking salient objects with their eyes. Here the assumption is that these patients do experience something, however minimal, at least some of the time.

The need to monitor consciousness also arises in a second, totally different set of patients who have a normally functioning brain—people who undergo invasive surgery for the usual host of ills, such as injuries, removal of a cancerous growth, or repair of knees, hips and other body parts. Anesthesia eliminates pain and other conscious experiences, prevents mobility and stabilizes the autonomic nervous system, which controls breathing and other functions, for hours at a time.

Patients "go under" with the expectation that they will not wake up during surgery and that they will not have to contend with traumatic memories of intraoperative experiences that could haunt them for the rest of their lives. Unfortunately, this goal is not always met. Intraoperative recall, or "awareness under anesthesia," can occur in a small number of operations, estimated to be in the one-per-1,000 range, in particular when patients are paralyzed during a procedure by an anesthesiologist to facilitate intubation and prevent gross muscle movements. Given that millions of Americans undergo surgical-level anesthesia every year, this tiny fraction translates into thousands of awakenings under anesthesia.

Existing EEG measures monitor depth of anesthesia during an operation. Yet none of the vast diversity of anesthetic agents work in a consistent manner across all patients, who range from neonates to birthing mothers, the very elderly or the very sick. What is needed is a tool that can reliably track the presence of consciousness in individual subjects across a large spectrum of normal and pathological conditions under both acute (anesthesia) and chronic conditions (the plight of neurologically impaired patients).

THE NATURE OF CONSCIOUS EXPERIENCE

TO DETECT consciousness, it is necessary to consider two essential characters of any subjective experience, no matter how mundane or exalted. First, by definition, any experience is different from all other experiences. It is specific to the moment and place in which it occurs. Each one is highly informative—take the unique visual richness associated with a mountain hike in the Rockies or another in the Cascade Range. Now combine these recollections with other sensory modalities, such as sounds and smells, emotions and memories. Each one is distinct in its own way. The second point is that each experience is seamless, integrated and holistic. You cannot separate the iconic percept of black smoke arising from the burning Twin Towers on a backdrop of blue sky into a half experience of the North Tower and another half experience of the South Tower.

The current most promising scientific theory of consciousness, which encompasses both of these ideas, is Integrated Information Theory (IIT). Devised by Giulio Tononi, a psychiatrist and neuroscientist at the University of Wisconsin–Madison, IIT emphasizes the differentiated and integrated aspect of any subjective experience and postulates that the mechanism supporting conscious experience in the human brain's neocortex must likewise incorporate these two attributes. To probe the extent to which these mechanisms are intact, Tononi, together with a team that included neurologist and neuroscientist Marcello Massimini, now at the University of Milan in Italy, devised an EEG-based method back in the early 2000s. It provides a very crude approximation of IIT's formal calculus. The team verified its basic soundness by correctly discriminating between when six healthy volunteers were conscious but quietly resting with eyes closed and when they were deeply asleep and therefore unconscious.

The brain of a deep sleeper acts like a stunted, badly tuned bell. Whereas the initial amplitude of the EEG is larger than when the subject is awake, its duration is much shorter, and it does not reverberate across the cortex to connected regions. While neurons remain active in deep sleep, as evidenced by the strong response in a local brain region, integration has broken down. Little of the electrical activity found in an awake brain is present.

Although distinguishing the brain's response during a restful state from its response while sleeping may seem trivial, the method can be extended to the more difficult task of discriminating among a variety of brain states. Indeed, in subsequent years, Tononi, Massimini and 17 additional doctors and brain scientists tested the procedure in many more subjects. A paper summarizing this landmark study was published in 2016 in the peer-reviewed literature.

The method perturbs the brain by sending one or two pulses of magnetic energy via an enclosed coil of wire held against the scalp, a method called transcranial magnetic stimulation, or TMS. This technique induces a brief electric current in the underlying cortical neurons, which, in turn, engage other neurons in a cascade that reverberates inside the head before the electrical surge dies out in a fraction of a second. Think of the brain as a large church bell and the TMS device as the clapper.

Once struck, a well-cast bell will ring at its characteristic pitch for a considerable time, as does the brain. Its electrical activity is monitored by a high-density EEG cap worn by the patient. The EEG is averaged and displayed during the course of 200 TMS pulses, as if it were a movie unfolding in time.

In an awake brain, with intact connectivity, this monitoring of different areas in response to the probe shows a highly complex pattern over much of the cortex, activity that is neither totally predictable nor completely random—and is emblematic of what is meant by "complex."

The researchers estimate its complexity, the extent to which this response differs across the cortex and across time, using a mathematical measure capturing its diversity. The technique itself is borrowed from computer science and is the basis of the popular "zip" compression algorithm for reducing the storage demand of images or movies, which is why the entire procedure of measuring consciousness is known in the trade as zap and zip. Ultimately each person's EEG response is mapped onto a single number, the perturbational complexity index, or PCI. If the brain does not react to the magnetic response—say, because the cortical activity is suppressed or only wiggles minimally—PCI will be close to zero, whereas maximal complexity yields a PCI of one. The larger the PCI, the more diverse the brain's response to the magnetic pulse.

ZAP AND ZIP IN PATIENTS

THE LOGIC OF the 2016 study, which involved patients from specialized clinics in Belgium and Italy, is straightforward. In a first step, zap and zip is applied to a control population to infer a cutoff value—tagged PCI*—above which consciousness is thought to be present. In every case in which consciousness can be reliably established in any one subject, the person's PCI values should be greater than PCI*, and in every case in which the subject is unconscious, PCI values should be below this threshold. This procedure establishes PCI* as a critical threshold—the minimum measure of complex brain activity—supporting consciousness. Then, in a second step, this threshold is used to infer whether consciousness is present in patients living in the gray zone, where more conventional measurements are insufficient.

In the study, the benchmark population used to calibrate the procedure encompassed two groups. One consisted of 102 volunteers with no known brain impairment who experienced various conscious or unconscious states: quietly awake with eyes closed or dreaming during REM sleep (the latter is also a conscious state the researchers assessed by randomly awakening the sleepers during REM sleep and only including their EEGs in the final results if they reported any dream experience immediately prior to awakening). The EEGs were also assessed under anesthesia using ketamine, a pharmacological agent that disconnects and dissociates the mind from the external world but does not extinguish consciousness. (At a lower dose, ketamine is abused as a hallucinogenic drug, known as vitamin K.) The unconscious conditions for which EEG was measured during the study were deep sleep (reporting no experiences immediately prior to being awakened) and surgical-level anesthesia using three different agents (midazolam, xenon and propofol). The study also included 48 brain-injured but responsive and awake patients who were assessed while awake as controls.

A measurement value derived by researchers enabled them to establish a critical threshold—the minimum degree of complex brain activity supporting consciousness.

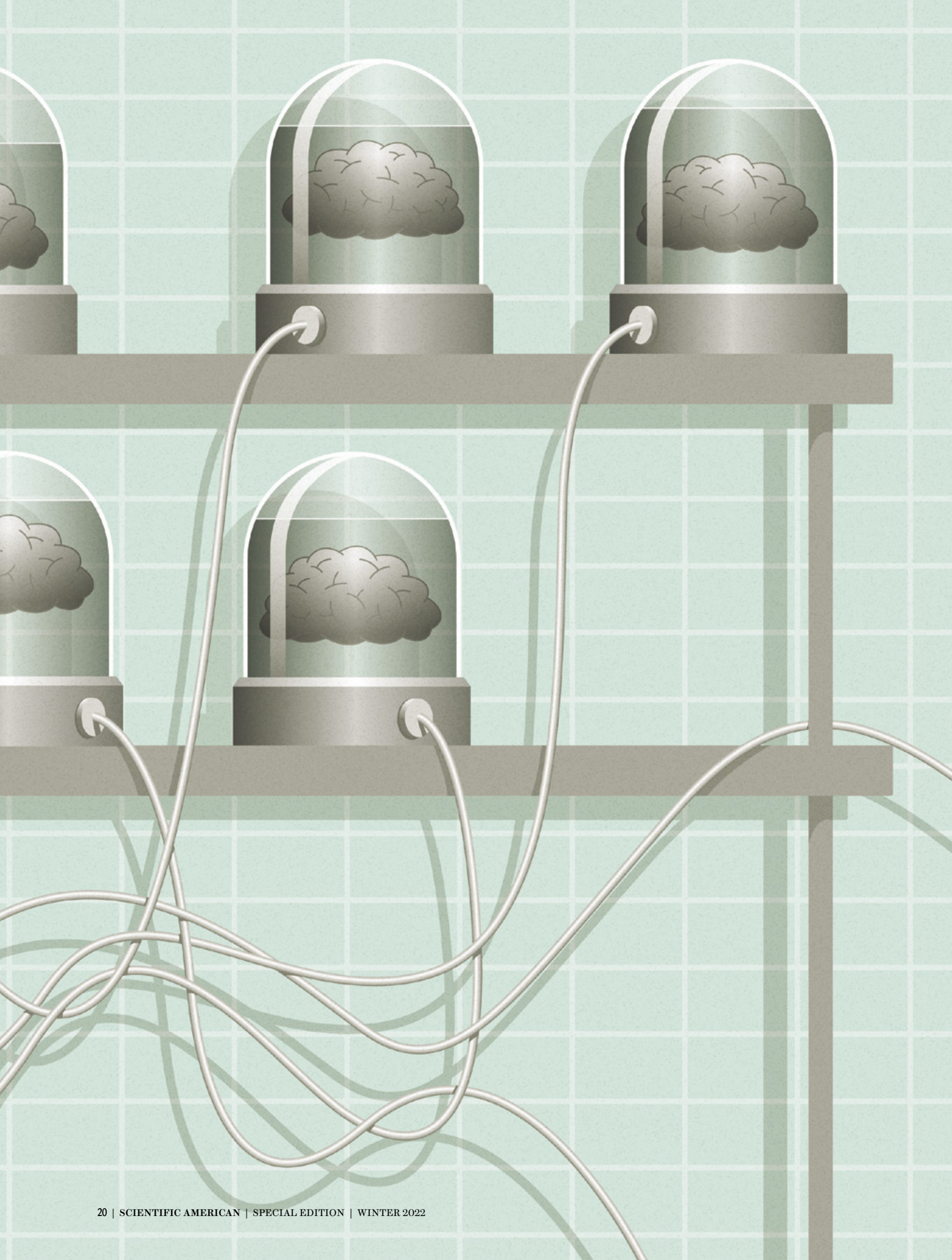
The investigators found that consciousness could be inferred with complete accuracy in every single subject using the same PCI* value of 0.31. That is, in every one of the 540 conditions tested across the 150 subjects, if the electrical response was at or below this threshold, the subject was unconscious. If above PCI*, the subject was conscious. Everyone in the study, whether an uninjured volunteer or a brain-injured patient, received a correct classification. This achievement is remarkable given the variability in gender, age, brain locations where the TMS pulses were applied, and medical and behavioral conditions in the study cohort.

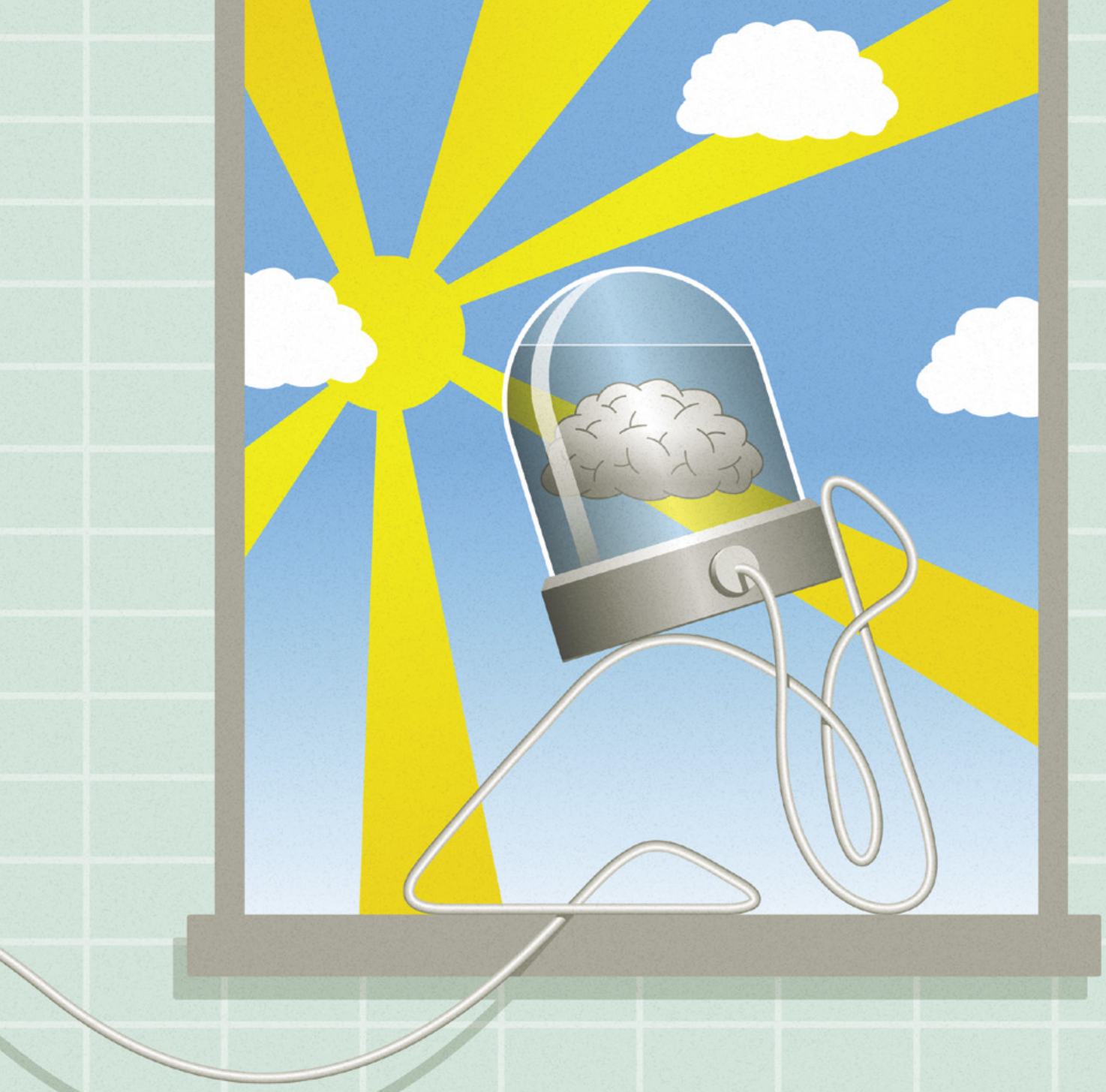
The team then applied zap and zip with this threshold value (of 0.31) to a distinct set of patients with severe disorders of consciousness—those either in a minimally conscious state or in an unresponsive wakeful one. In the MCS group, consisting of patients with at least some signs of behavior beyond reflexive functions such as breathing, the method correctly assigned consciousness to 36 of 38 patients, misdiagnosing the other two as unconscious. Of the 43 UWS patients, in which communication failed, 34 had a brain response whose complexity was less than that of anyone of the benchmark population when conscious, an expected result. That is, the complexity of their EEG responses was comparable to that of the benchmark group when not detecting consciousness.

Much more troubling, however, were the other nine patients who responded to the TMS pulse with a complex pattern of electrical activity that lies above the threshold. That is, the perturbational complexity of their brains' responses was as high as in many conscious benchmark controls. These patients with high-complexity cortical responses may experience something yet are unable to communicate with the world and their loved ones.

As any successful experiment does, this one is leading to additional clinical studies. How can the zap-and-zip method be improved to achieve 100 percent accuracy in minimally conscious patients? Could other groups of patients, such as those with catatonia or late-stage dementia, infants, or young children, also be tested? Another question is whether other physiological or behavioral measures can be developed to corroborate the inference that some UWS patients are conscious. Can the method be turned into a prognostic tool, inferring to what extent UWS patients are on the road to recovery? Those questions need to be tackled moving forward. But in the interim, let us celebrate a milestone in untangling the ancient mind-body problem. SA

Christof Koch is chief scientist of MindScope at the Allen Institute for Brain Science and of the Tiny Blue Dot Foundation, as well as author of *The Feeling of Life Itself—Why Consciousness Is Widespread but Can't Be Computed* (MIT Press, 2019). He is on *Scientific American*'s board of advisers.





Can Lab-Grown Brains Become Conscious?

A handful of experiments are raising questions about whether clumps of cells and disembodied brains could be sentient and how scientists would know if they were • *By Sara Reardon*

Illustration by Thomas Fuchs

IN ALYSSON MUOTRI'S LABORATORY, HUNDREDS OF MINIATURE HUMAN BRAINS, the size of sesame seeds, float in petri dishes, sparking with electrical activity. These tiny structures, known as brain organoids, are grown from human stem cells and have become a familiar fixture in many labs that study the properties of the brain. Muotri, a neuroscientist at the University of California, San Diego, has found some unusual ways to deploy his. He has connected organoids to walking robots, modified their genomes with Neandertal genes, launched them into orbit onboard the International Space Station and used them as models to develop more humanlike artificial-intelligence systems. Like many scientists, Muotri temporarily pivoted to studying COVID, using brain organoids to test how drugs perform against the SARS-CoV-2 coronavirus.

But one experiment has drawn more scrutiny than the others. In August 2019 Muotri's group published a paper in *Cell Stem Cell* reporting the creation of human brain organoids that produced coordinated waves of activity resembling those seen in premature babies. The waves continued for months before the team shut the experiment down.

This type of brain-wide, coordinated electrical activity is one of the properties of a conscious brain. The team's finding led ethicists and scientists to raise a host of moral and philosophical questions about whether organoids should be allowed to reach this level of advanced development, whether "conscious" organoids might be entitled to special treatment and rights not afforded to other clumps of cells and the possibility that consciousness could be created from scratch.

The idea of bodiless, self-aware brains was already on the minds of many neuroscientists and bioethicists. Just a few months earlier a team at Yale University announced that it had at least partially restored life to the brains of pigs that had been killed hours before. By removing the brains from the pigs' skulls and infusing them with a chemical cocktail, the researchers revived the neurons' cellular functions and their ability to transmit electrical signals.

Other experiments, such as efforts to add human neurons to mouse brains, have raised questions, with some scientists and ethicists arguing that these experiments should not be allowed.

The studies set the stage for a debate between those who want to avoid the creation of consciousness and those who see complex organoids as a means to study and test treatments for human diseases. Muotri and many other neuroscientists think that human brain organoids could be the key to understanding uniquely human conditions such as autism and schizophrenia, which are impossible to study in detail in mouse models. To achieve this goal, Muotri says, he and others might need to deliberately create consciousness.

Researchers have been calling for a set of guidelines, similar to those used in animal research, to guide the humane use of brain organoids and other experiments that could achieve consciousness. In June 2020 the U.S. National Academies of Sciences, Engineering and Medicine began a study with the aim of outlining the potential legal and ethical issues associated with brain organoids and human–animal chimeras.

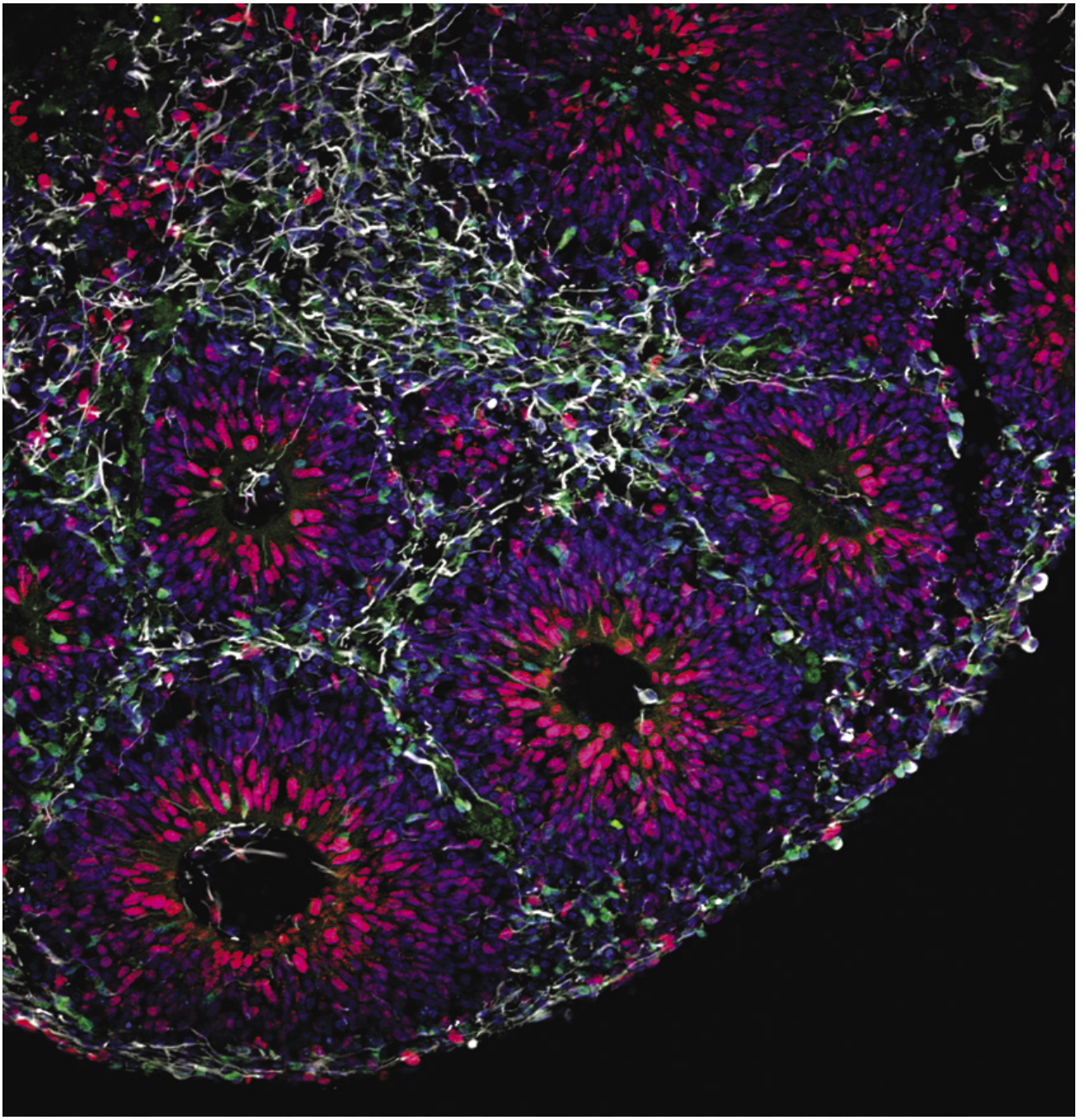
The concerns over lab-grown brains have also highlighted a problem: neuroscientists have no agreed way to define and measure consciousness. Without a working definition, ethicists worry that it will be impossible to stop an experiment before it crosses a line.

The current crop of experiments could force the issue. If scientists become convinced that an organoid has gained consciousness, they might need to hurry up and agree on a theory of how that happened, says Anil K. Seth, a cognitive neuroscientist at the University of Sussex in England. But, he says, if one person's favored theory deems the organoid conscious, whereas another's does not, any confidence that consciousness has been attained vanishes. "Confidence largely depends on what theory we believe in. It's a circularity."

SENTIENT STATES

CREATING A CONSCIOUS SYSTEM might be a whole lot easier than defining it. Researchers and clinicians define consciousness in many different ways for various purposes, but it is hard to synthesize them into one neat operational definition that could be used to decide on the status of a lab-grown brain.

Physicians generally assess the level of consciousness in patients in a vegetative state on the basis of whether the person blinks or flinches in response to pain or other stimuli. Using electroencephalogram (EEG) readings, for instance, researchers can also measure how the brain responds when it is zapped with an electrical pulse. A conscious brain will display much more complex, unpredictable electrical activity than one that is unconscious, which responds with simple, regular patterns.



IN DEVELOPING human brain organoids, preneuronal cells (*red*) turn into neurons (*green*), which wire up into networks (*white*).

Muotri Lab/University of California, San Diego

But such tests might not adequately probe whether a person lacks consciousness. In brain-imaging studies of people who are in a coma or a vegetative state, scientists have shown that unresponsive individuals can display some brain activity reminiscent of consciousness—such as activity in motor areas when asked to think about walking.

In any case, standard medical tests for consciousness are difficult to apply to brain cells grown in dishes or to disembodied animal brains. When Muotri suggested that his organoids' firing patterns were just as complex as those seen in preterm infants, people were unsure what to make of that. Some researchers do not consider the brain activity in a preterm infant to be complex enough to be classed as conscious. And organoids cannot blink or recoil from a painful stimulus, so they would not pass the clinical test for consciousness.

In contrast, it is much more likely that an intact brain from a recently killed pig has the necessary structures for consciousness, as well as wiring created by memories and experiences the animal had while it was alive. "Thinking about a brain that has been filled with all this, it is hard to imagine that brain would be empty," says Jeantine Lunshof, a philosopher and neuroethicist at Harvard University. "What they can do in terms of thinking, I don't know, but it's for sure not zero," Lunshof says. Bringing a dead brain back to a semblance of life, as the Yale team did,

might have the potential to restore a degree of consciousness, although the scientists took pains to avoid this by using chemical blocking agents that prevented brain-wide activity.

Researchers agree that they need to take the possibilities raised by these studies seriously. In October 2019 U.C.S.D. held a conference of about a dozen neuroscientists and philosophers, together with students and members of the public, with the intention of establishing and publishing an ethical framework for future experiments. But the paper was delayed for months, partly because several of the authors could not agree on the basic requirements for consciousness.

INCREASINGLY COMPLEX

SO FAR NOBODY has created consciousness in the lab, say scientists and ethicists who study the issue. But they are asking themselves what to watch out for and which theories of consciousness might be most relevant. According to an idea called integrated information theory, for example, consciousness is a product of how densely neuronal networks are connected across the brain. The more neurons that interact with one another, the higher the degree of consciousness—a quantity known as phi. If phi is greater than zero, the organism is considered conscious.

Most animals reach this bar, according to the theory. Christof Koch, who serves on *Scientific American*'s board of advisers and is chief scientist of the MindScope Program at the Allen Institute for Brain Science, doubts any existing organoid could achieve this threshold but concedes that a more advanced one might.

Other competing theories of consciousness require sensory input or coordinated electrical patterns across multiple brain regions. An idea known as global workspace theory, for instance, posits that the brain's prefrontal cortex functions as a computer, processing sensory inputs and interpreting them to form a sense of being. Because organoids do not have a prefrontal cortex and cannot receive input, they cannot become conscious. "Without input and output, the neurons may be talking with each other, but that doesn't necessarily mean anything like human thought," says Madeline Lancaster, a developmental biologist at the University of Cambridge.

Connecting organoids to organs, however, could be a fairly simple task. In 2019 Lancaster's team grew human brain organoids next to a mouse spinal column and back muscle. When nerves from the human organoid connected with the spinal column, the muscles began to spontaneously contract.

Most organoids are built to reproduce only one portion of the brain—the cortex. But if they develop long enough and with the right kinds of growth factor, human stem cells spontaneously re-create many different parts of the brain, which then begin coordinating their electrical activity. In a study published in 2017, molecular biologist Paola Arlotta of Harvard coaxed stem cells to develop into brain organoids composed of many different cell types, including light-sensitive cells like those found in the retina. When exposed to light, neurons in the organoids began firing. But the fact that these cells were active does not mean the organoids could see and process visual information, Arlotta says. It simply means that they could form the necessary circuits.

Arlotta and Lancaster think their organoids are too primitive to be conscious because they lack the anatomical structures necessary to create complex EEG patterns. Still, Lancaster admits that for advanced organoids, it depends on the definition. "If you thought a fly was conscious, it's conceivable that an organoid could be," she says.

Lancaster and most other researchers think that something like a revitalized pig brain would be much more likely to achieve consciousness than an organoid. The team that did the work on the pig brains, led by neuroscientist Nenad Sestan, was trying to find new ways to revitalize organs, not to create consciousness. The researchers were able to get individual neurons or groups of them to fire and were careful to try to avoid the creation of widespread brain waves. Still, when Sestan's team members saw what looked like coordinated EEG activity in one of the brains, they immediately halted the project. Even after a neurology specialist confirmed that the pattern was not consistent with consciousness, the group anesthetized the brains as a precautionary measure.

Sestan also contacted the U.S. National Institutes of Health for guidance on how to proceed. The agency's neuroethics panel, including Lunshof and Insoo Hyun, a bioethicist at Case Western University, assessed the work and agreed that Sestan should continue to anesthetize the brains. But the panel has not settled on more general regulations and does not routinely require a bioethics assessment for organoid proposals, because its members think that consciousness is unlikely to arise. The NIH has not arrived at a definition of consciousness, either. "It's so flexible, everyone claims their own meaning," Hyun says. "If it's not clear we're talking about the same thing, it's a big problem for discourse."

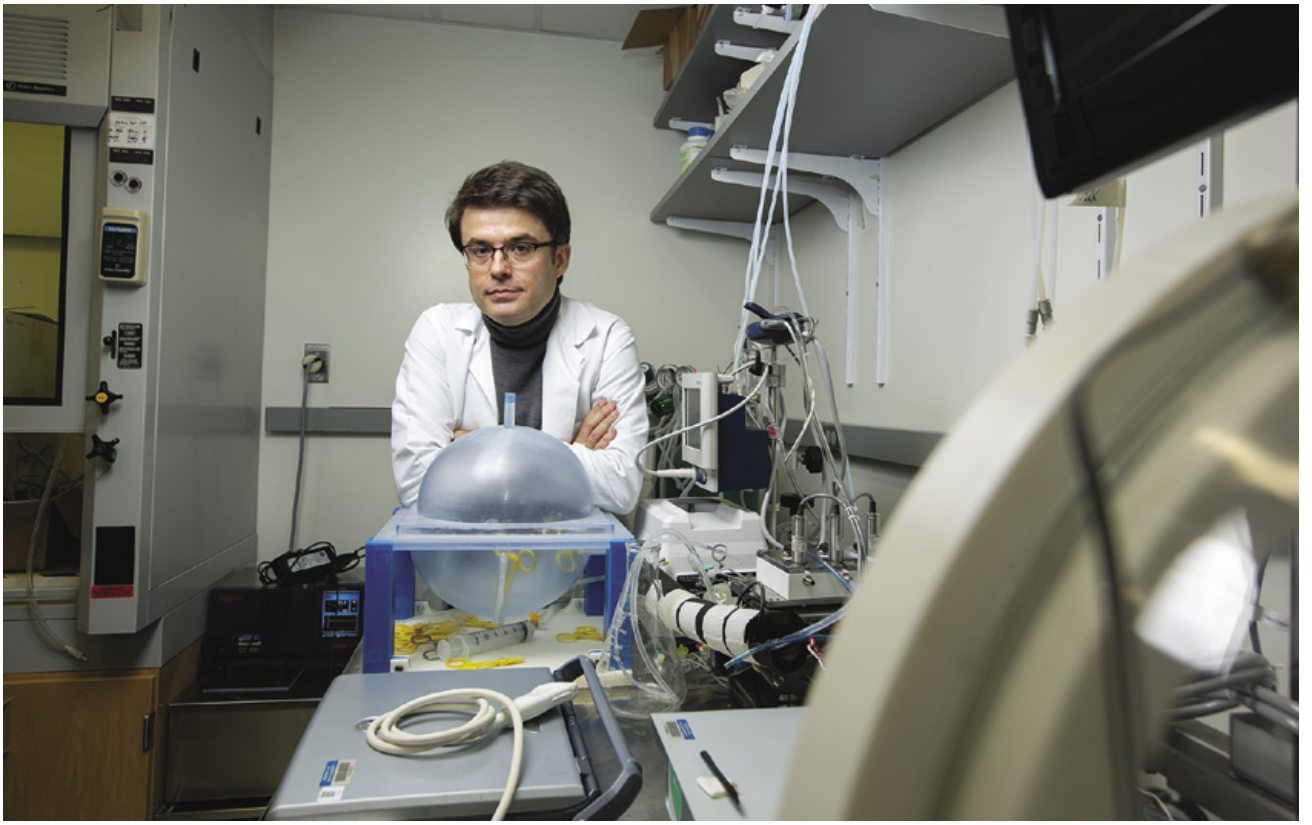
FUZZY DEFINITIONS

SOME THINK IT IS FUTILE to even try to identify consciousness in any sort of lab-maintained brain. "It's just impossible to say meaningful things about what these bunches of brain cells could think or perceive, given we don't understand consciousness," says Steven Laureys, a neurologist at the University of Liège in Belgium, who pioneered some of the imaging-based measures of consciousness in people in a vegetative state. "We shouldn't be too arrogant." Further research should proceed very carefully, he says.

Laureys and others point out that the experience of an organoid is likely to be very different from that of a preterm infant, an adult human or a pig and would not be directly comparable. Furthermore, the structures in an organoid might be too small to have their activity measured accurately, and similarities between the EEG patterns of organoids and of preterm baby brains could be coincidental. Other scientists who work on brain organoids also caution against making assumptions about the link between activity patterns in organoids and consciousness.

"This system is not the human brain," says Sergiu Pasca, a neuroscientist at Stanford University. "They're made out of neurons. Neurons have electrical activity, but we have to think carefully about how to compare them."

Muotri wants his organoid systems to be comparable, in at least some ways, with human brains so that he can study human disorders and find treatments. His motivation is personal: his teenage son has epilepsy and autism. "He struggles hard in life," Muotri says. Brain organoids are a promising avenue because they recapitulate the earliest stages of brain wiring, which are impossible to study as a human embryo develops. But studying human brain disorders without a fully func-



NEUROSCIENTIST Nenad Sestan used the BrainEx platform to restore neural activity in disembodied pig brains.

Jesse Winter

tioning brain, he says, is like studying a pancreas that does not produce insulin. "To get there, I need a brain organoid model that really resembles a human brain. I might need an organoid that becomes conscious."

Muotri says he is agnostic about which definition to use to decide whether an organoid reaches consciousness. At some point, he says, organoids might even be able to help researchers answer questions about how brains produce conscious states. For instance, mathematician Gabriel Silva of U.C.S.D. is studying neural activity in Muotri's organoids to develop an algorithm that describes how the brain generates consciousness. The goal of his project, which is partially funded by Microsoft, is to create an artificial system that works like human consciousness.

At the moment, there are no regulations in the U.S. or in Europe that would stop a researcher from creating consciousness. The National Academies panel released a report in April 2021 outlining the latest research and what it views as appropriate oversight. Members weighed in on questions such as whether to obtain people's consent to develop their cells into brain organoids and how to study and dispose of organoids humanely. The International Society for Stem Cell Research has also released organoid guidelines but is not addressing consciousness, because it does not think the science is there yet.

Hyun says that the NIH neuroethics panel has not yet seen any proposals to create complex, conscious organoids that would necessitate new guidelines. And Muotri says he does not know of anyone else deliberately trying to create conscious organoids either, although a sufficiently complex organoid could, by some definitions, reach that status accidentally.

Still, Muotri and others say they would welcome some stricter guidelines. These could include requiring scientists to justify the number of human brain organoids they use, to use them only for research that cannot be done in any other way, to restrict the amount of pain that can be inflicted on them, and to dispose of them humanely.

Having such advice in place ahead of time would help researchers weigh the costs and benefits of creating conscious entities. And many researchers emphasize that such experiments have the potential to yield important insights. "There are truly conscious people out there with neurological disorders with no treatments," Lancaster says. "If we did stop all of this research because of the philosophical thought experiment," she adds, "that would be very detrimental to actual human beings who do need some new treatment."

Treatments could still, however, be tested in brain organoids made using mouse stem cells or in regular animal models. Such experiments could also inform discussions about the ethical use of human organoids. For instance, Hyun would like to see researchers compare the EEG patterns of mouse brain organoids with those of living mice, which might indicate how well human organoids recapitulate the human brain.

For his part, Muotri sees little difference between working on a human organoid and using a lab mouse. "We work with animal models that are conscious, and there are no problems," he says. "We need to move forward, and if it turns out they become conscious, to be honest, I don't see it as a big deal." SA

Sara Reardon is a freelance reporter based in Bozeman, Mont.

The Brain Electric

Electrodes that stimulate brain tissue reveal the topography of conscious experience

By Christof Koch

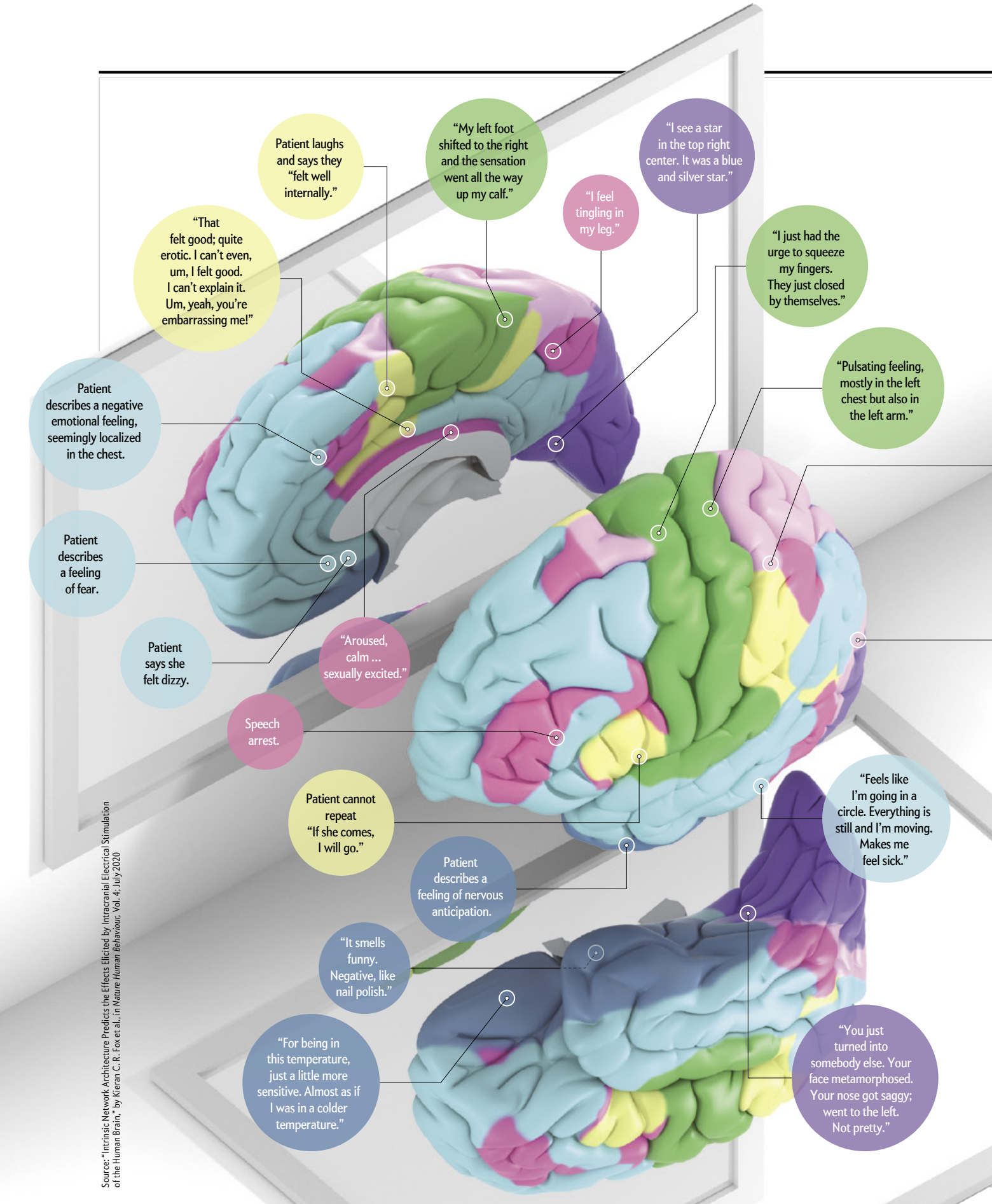
Illustration by Zara Picken

Consider the following experiences:

- You're headed toward a storm that's a couple of miles away, and you've got to get across a hill. You ask yourself: "How am I going to get over that, through that?"
- You see little white dots on a black background, as if looking up at the stars at night.
- You look down at yourself lying in bed from above but see only your legs and lower trunk.

These may seem like idiosyncratic events drawn from the vast universe of perceptions, sensations, memories, thoughts and dreams that make up our daily stream of consciousness. In fact, each one was evoked by directly stimulating the brain with an electrode. As American poet Walt Whitman intuited in his poem "I Sing the Body Electric," these anecdotes illustrate the intimate relationship between the body and its animating soul. The brain and the conscious mind are as inexorably linked as the two sides of a coin.

Recent clinical studies have uncovered some of the laws and regularities of conscious activity, findings that have occasionally proved to be paradoxical. They show that brain areas involved in conscious perception have little to do with thinking, planning and other higher cognitive functions. Neuroengineers are now working to turn these insights into technologies to replace lost


Patient laughs and says they "felt well internally."
"My left foot shifted to the right and the sensation went all the way up my calf."
"I see a star in the top right center. It was a blue and silver star."
"I feel tingling in my leg."
"That felt good; quite erotic. I can't even, um, I felt good. I can't explain it. Um, yeah, you're embarrassing me!"
"I just had the urge to squeeze my fingers. They just closed by themselves."
Patient describes a negative emotional feeling, seemingly localized in the chest.
"Pulsating feeling, mostly in the left chest but also in the left arm."
Patient describes a feeling of fear.
Patient says she felt dizzy.
"Aroused, calm ... sexually excited."
Speech arrest.
Patient cannot repeat "If she comes, I will go."
"Feels like I'm going in a circle. Everything is still and I'm moving. Makes me feel sick."
Patient describes a feeling of nervous anticipation.
"It smells funny. Negative, like nail polish."
"For being in this temperature, just a little more sensitive. Almost as if I was in a colder temperature."
"You just turned into somebody else. Your face metamorphosed. Your nose got saggy; went to the left. Not pretty."

Source: "Intrinsic Network Architecture Predicts the Effects Elicited by Intracranial Electrical Stimulation of the Human Brain," by Kieran C. R. Fox et al., in *Nature Human Behaviour*, Vol. 4, July 2020

Where Experiences Live in the Brain

An atlas published in 2020 compiled the verbal reports of people with epilepsy whose cortical areas were stimulated with electrodes during surgery. What they felt and perceived varied depending on which brain region was stimulated. All of the 1,537 locations in these 67 patients where current was applied were mapped onto a digital brain model, a simplified version of which is depicted here. When stimulated at these sites, patients talked about their experiences.

"Just really couldn't move (my fingers) too much; lost the motion. The hand felt a little tight, but the thumb was out of commission."

"I felt like my arms were moving but they weren't. I felt side-to-side movements, like floating in the air."

Colors of Cognition

Stimulation was applied to varied brain networks and regions.

- Somatomotor
- Visual
- Dorsal attention
- Salience
- Frontoparietal
- Limbic
- Default

Points of Stimulation

Researchers placed dozens of electrodes onto the surface of the cerebral cortex accessed through an opening in the skull.

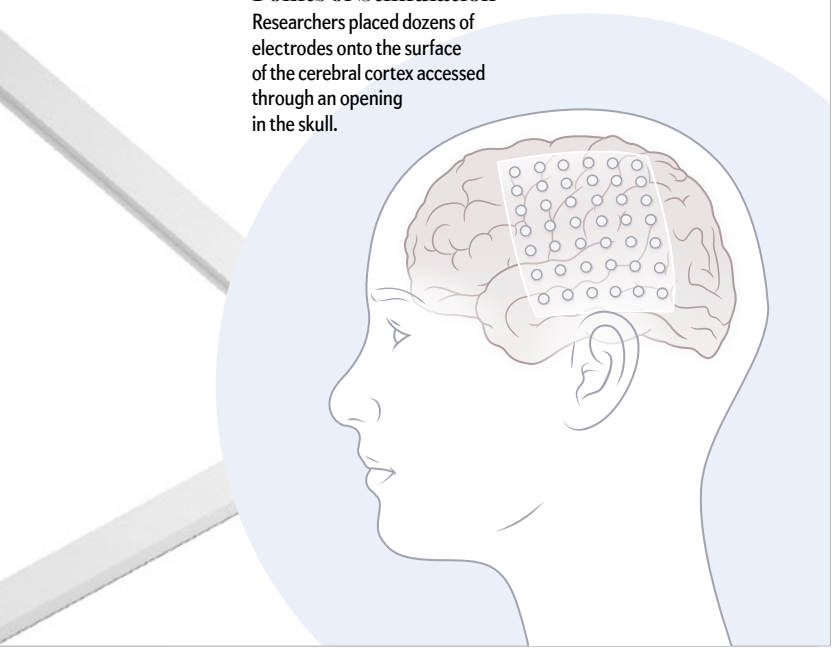


Illustration by AXS Biomedical Animation Studio

cognitive function and, in the more distant future, to enhance sensory, cognitive or memory capacities. For example, a recent brain-machine interface provides completely blind people with limited abilities to perceive light. These tools, however, also reveal the difficulties of fully restoring sight or hearing. They underline even more the snags that stand in the way of sci-fi-like enhancements that would enable access to the brain as if it were a computer storage drive.

ANIMAL ELECTRICITY

NERVOUS SYSTEMS operate on the flow of electric currents through ultradense and hyperconnected networks of switching elements. Countless physicians and scientists have worked on this problem over the past two and a half centuries, beginning with Italian physician Luigi Galvani, who in the late 18th century connected a freshly killed frog to a long metal wire. By pointing the wire toward the sky during a thunderstorm, he made the frog's leg jump and twitch with each flash of lightning. Galvani's investigations revealed that nerve fibers transmitted "animal electricity," which is no different in kind from the "atmospheric electricity" that Benjamin Franklin demonstrated with his kite experiments in Philadelphia in 1752. In 1802 Galvani's nephew Giovanni Aldini electrically stimulated the exposed brain of a decapitated prisoner during a public event. A jaw quivered. An eye opened. The spectacle may have helped to inspire Mary Shelley to write the classic gothic novel *Frankenstein*.

Subsequent animal studies demonstrated that exciting particular brain regions triggered movements in specific muscles and limbs. These investigations led to the discovery of the motor cortex in the 1870s. In 1874 American physician Robert Bartholow performed the first direct brain stimulation of a conscious patient—a pioneering act clouded in ethical controversy because it caused the patient pain and probably hastened her death. Intracranial electrical stimulation (iES) was refined over the following decades. It became part of the neurosurgeon's toolbox thanks to the ground-breaking work of Wilder Penfield of the Montreal Neurological Institute, who between the 1930s and the 1950s used iES to map cortical areas that process motor or sensory functions.

In some people with epilepsy, drugs fail to adequately control the number or severity of seizures. Neurosurgery becomes an option if those seizures originate in a delimited neighborhood in the cortex—the outermost layer of the brain involved in perception, motor control, speech, reasoning, and so on—or in closely related structures, such as the hippocampus. Uncontrolled hyperexcitability starts because of local faulty wiring. It can grow and eventually engulf the rest of the brain. How much tissue to remove is a dilemma: cut too little, and seizures may continue; cut too much, and the patient may lose the ability to speak, see or walk. Surgeons must avoid areas of the cortex that are crucial for everyday behavior, such as the primary auditory, visual, somatosensory and motor cortices and the regions controlling understanding and producing speech, areas known as the eloquent cortex.

iES is brought in as a means to look for tissue that needs preserving. Neurosurgeons implant disk-shaped electrodes inside the skull, underneath the tough, leatherlike membrane known as the dura mater. Alternatively, they may insert needlelike electrodes into the brain's gray matter to probe its function. Once the surgeons have identified the focal point of the seizure and removed the electrodes, they cut or coagulate this tissue in a follow-up operation, and the patient usually becomes seizure-free.

A different use for iES is chronic electrical stimulation, in which the electrodes are left permanently in place. Gentle pulses of current sent through the electrodes can control the tremors and rigidity of Parkinson's disease (a technique called deep-brain stimulation) or reduce the incidence and severity of seizures. Pilot clinical experiments are evaluating the use of such implanted electrodes as a visual prosthetic device to enable people with vision impairments to navigate and as a therapy for obsessive-compulsive disorder and depression.

Patients reported electrode-evoked experiences such as seeing distorted faces reminiscent of paintings by Salvador Dalí.

HOT OR NOT

IN JULY 2020 *Nature Human Behaviour* published an atlas highlighting locations across the cortex that, when aroused with electrodes, evoked conscious experiences, such as the storm and the disconnected body mentioned earlier. Led by Josef Parvizi, a professor of neurology at the Stanford University School of Medicine, the clinical team collected data from 67 people with epilepsy. The researchers recorded electrical activity from more than 1,500 sites in the cortex, primarily with subdural electrodes. They mapped the recordings from those sites to spots on a digital brain model so they could compare data from different brains (the pattern of ridges and valleys that give the organ the look of an oversized walnut differs from person to person). The team looked for "responsive" electrodes that triggered some visual or tactile sensation, muscle twitching or disrupted speech. If the patient did not feel anything when stimulated, that electrode was marked as nonresponsive.

Patients reported a range of electrode-evoked subjective experiences: briefly flashing points akin to stars of light; distorted faces like those in the paintings of Salvador Dalí; bodily feelings such as tingling, tickling, burning, pulsing and so-called out-of-body experiences; fear, unease, sexual arousal, merriment; the desire to move a limb; the will to persevere in the face of some great but unrecognized challenge. Mere tickling of neural tissue with a tiny bit of electric current was enough to evoke these feelings. During sham stimulation (no current applied), patients did not feel anything.

Although iES is safe and effective, it is also crude. The low-impedance electrodes are six to 10 square millimeters in area and deliver up to 10 milliamperes of electric current between adjacent electrodes—enough to modulate the excitability of a million or more nerve cells. Still, effects induced by iES can be quite localized. Responsiveness can change from all to none within millimeters or across a sulcus (a groove on the cortical surface).

The Parvizi team found that electrodes in the dedicated sensory and motor areas were far more likely to be responsive than those in areas of the cortex that process higher cognitive functions. Half to two thirds of electrodes above visual and tactile (somatosensory) cortex areas triggered some conscious perception; in regions of the lateral and anteromedial prefrontal cortex, which are involved with higher thought processes, at most one in five electrodes did so. Put differently, electrodes in the back of the cortex—in areas responsible for sensory experiences—were more likely to be active than those toward the front, which consists of regions of the cortex important for cognitive activity such as thinking, planning, moral reasoning, decision-making and intelligence.

Despite their importance for thinking, these regions have little to do with consciousness. Indeed, for the past century neurosurgeons have observed that so long as the eloquent cortex is spared, massive regions of the prefrontal cortex can be ablated without causing obvious deficits in the daily stream of consciousness of these patients. These regions of noneloquent cortex can modulate consciousness, but they are, by and large, not where conscious experience appears to originate. That privilege belongs to more posterior regions—the parietal, temporal and occipital lobes. Why the physical substrate of our mental experiences should be in the back rather than in the front of the brain remains a mystery.

TO SEE OR NOT TO SEE

APPLYING iES to the visual cortex triggers optical sensations known as phosphenes, brief flashes that resemble lightning striking a darkened plain. This observation is the source of a long-standing dream of a prosthetic device that restores some vision to people who are blind. Millions worldwide live with deficits in both eyes from retinitis pigmentosa, age-related macular degeneration, glaucoma, infection, cancer or trauma.

Doctors, scientists and engineers started pursuing visual prosthetics in the 1960s but have only recently been able to harness the appropriate technology to help blind people. One prominent example is a device known as Orion, developed by Second Sight Medical Products in Los Angeles. A tiny camera, mounted on glasses, converts images into pulses and transmits them wirelessly to fire 60 electrodes sitting on the visual cortex. The handful of people who have had this experimental device implanted into their brain perceive clouds of dots that allow them to navigate. "It's still a blast every time I turn it on," one study participant reports. "After seeing noth-

ing to all of a sudden seeing little flickers of light move around and figuring out that they mean something. It's just amazing to have some form of functional vision again." Orion significantly improves the quality of life for people who previously lived in complete darkness. It enables them to safely cross the street or locate a doorway. But it does not allow them to regain the ability to recognize figures, shapes or letters.

A team at the University of California, Los Angeles, and the Baylor College of Medicine led by neurosurgeon Daniel Yoshor recently did accomplish this feat, as described in the journal *Cell*. They stimulated nearby locations in the visual cortex to trigger phosphenes that appear close together, demonstrating that the external visual environment is mapped in a regular fashion onto the surface of the visual cortex. This observation has led to the erroneous belief that individual phosphenes are like pixels on a computer display—that is, if you were to simultaneously stimulate a series of points on the cortical surface in the shape of a cross, the subject should see points forming a cross. This does not happen, however.

Stimulating more than one location yields unpredictable results. In one participant, simultaneous stimulation of five electrodes, each one associated with one discrete phosphene, triggered the illumination of two large phosphenes that did not coalesce into a letter or any other coherent form. If the researcher staggered activation of the electrodes in time, however, the subject could identify shapes. The staggering reflected the delay required to trace the shape of a letter, as if the researcher were outlining a letter into the hand of the subject or onto a piece of paper. In this more dynamic manner, the subject with the implant whose vision was blocked could identify a stimulus by tracing out a Z, N, V and W, rapidly distinguishing upward from downward motion or discriminating sequences of letters.

Seeing the shape of a single letter is not quite the same as seeing a glorious sunset over Homer's wine-dark sea, but it represents progress. Why staggering stimulation in time improves perception is not clear and reveals our ignorance concerning functioning cortical circuits.

WHAT LIES AHEAD

TECHNOLOGICAL PROGRESS in so-called brain-machine interfaces is proceeding at a rapid pace. Elon Musk's company Neuralink released in April 2021 an impressive video showcasing a monkey playing the computer game Pong without any controller. This was achieved with two small chips implanted into the left and right motor cortices of the animal. Each chip has 1,024 hairlike electrodes that record the chattering of individual neurons. Collectively they convey the monkey's intention to quickly move the paddle up or down the screen to return the ball to the opposite side. Everything was done wirelessly; no electronics or dangling wires were protruding from the monkey's head. Many assume that surgeons will soon routinely replace or bypass faulty biological components—defective eyes or ears, failing memories—with superior electronic substitutes. Such optimism neglects the fact that all of this requires trepanation of the skull. In general, turning scientific insights into actual therapeutics is done in decades rather than in years. I am pretty confident that such enhancements will not occur in my lifetime (I'm now 65).

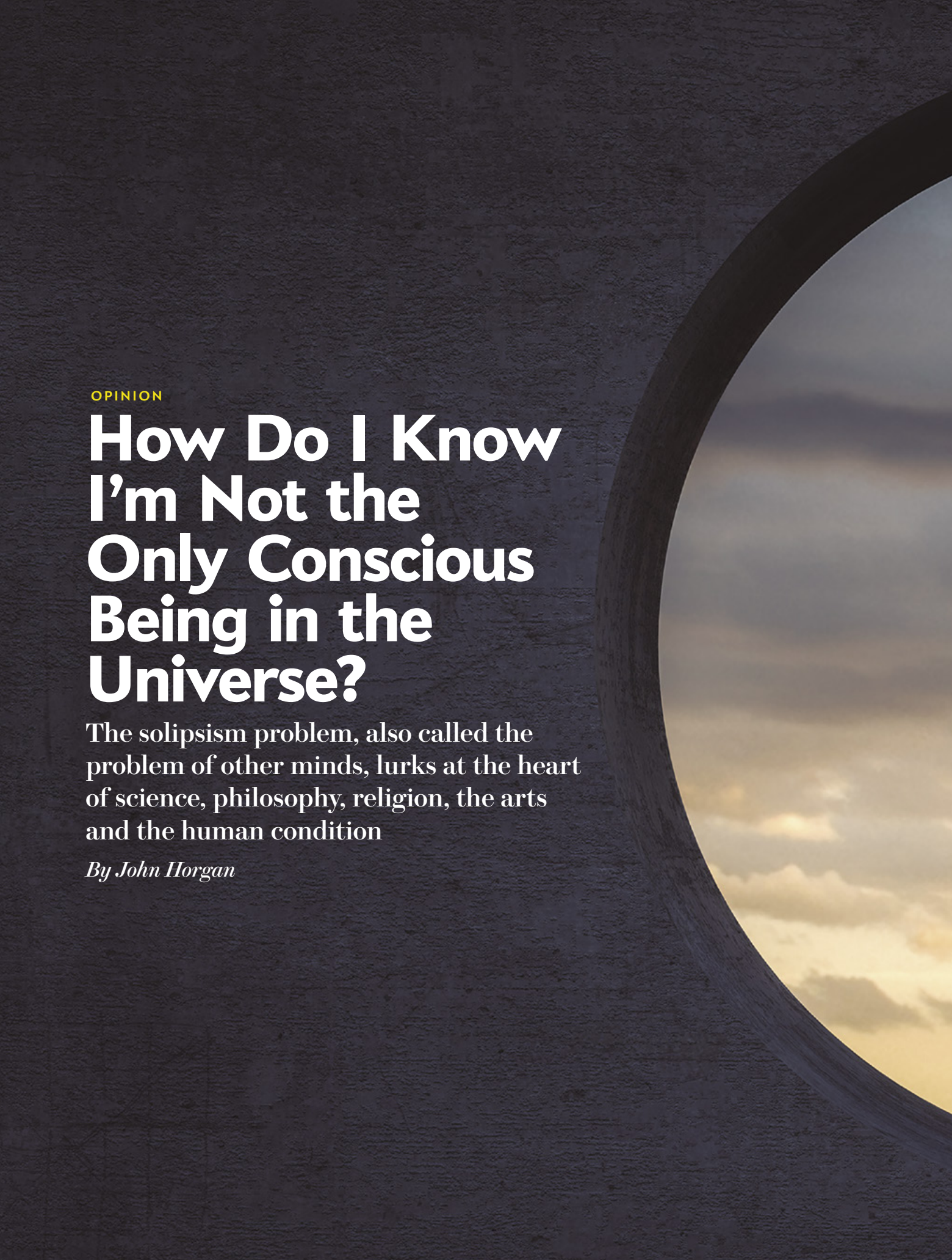
The "easiest" hurdles to overcome on the way to such a utopian (or perhaps dystopian) future are technological ones—reliably, quickly, and delicately reading and writing the brain electric. Neuralink's device represents the best of currently available technology and will certainly improve in future iterations. But we still have a long way to go before we can identify which of the 50,000 or more neurons in any quinoa-sized bit of brain matter are involved in any given perception or action. Only when that happens will it be possible to limit electrical stimulation to just those neurons and avoid stimulating nearby cells or output cables. That Parvizi and his colleagues failed to elicit conscious perceptions in more than half of all stimulated sites shows we lack tools capable of reliably eliciting *any* arbitrary sensation through electrical stimulation, let alone being able to evoke any highly specific one.

Even more challenging are surgical and regulatory hurdles that demand that prosthetic devices can be routinely and safely implanted by drilling through the hard skull into the gray matter underneath while minimizing the risk of infections, bleeding and seizures. Furthermore, the electronics has to function for years inside warm, wet and salty biological tissue—hardly an optimal operating regime. You don't want your prosthetic device to corrode or freeze up in the equivalent of the blue screen of death. For this reason, neural implants will remain a matter of last resort for those with severe sensory or motor impairments. As neuroprosthetic devices move through clinical trials, they will help people with visual impairments see and paralyzed patients with spinal cord damage to text or to steer a wheelchair with their thoughts, like the mind-Pong-playing monkey. For everyone else, the benefits of highly invasive brain surgery are unlikely to outweigh the costs.

But the true Annapurna ahead involves understanding how three pounds of excitable brain matter is responsible for seeing, moving and suffering. Yes, the physical substrate of heaven and hell is rooted in bioelectric signals that obey natural laws. But that tells us precious little about how a trillion electrical spiking signals each second, streaming over networks of tens of billions of heterogeneous cells, constitute a sight, sound or emotion.

Intracranial brain stimulation highlights the daily miracle of the brain's water changing into the wine of consciousness. The question remains, though: What is it about the brain, the most complex piece of active matter in the known universe, that turns the activity of 86 billion neurons into the feeling of life itself? SA

Christof Koch is chief scientist of MindScope at the Allen Institute for Brain Science and of the Tiny Blue Dot Foundation, as well as author of *The Feeling of Life Itself—Why Consciousness Is Widespread but Can't Be Computed* (MIT Press, 2019). He is on *Scientific American*'s board of advisers.



OPINION

How Do I Know I'm Not the Only Conscious Being in the Universe?

The solipsism problem, also called the problem of other minds, lurks at the heart of science, philosophy, religion, the arts and the human condition

By John Horgan

IT IS A CENTRAL DILEMMA OF HUMAN LIFE—MORE URGENT, ARGUABLY, than the inevitability of suffering and death. I have been brooding and ranting to my students about it for years. It surely troubles us more than ever during this plague-ridden era. Philosophers call it the problem of other minds. I prefer to call it the solipsism problem. Solipsism, technically, is an extreme form of skepticism, at once utterly illogical and irrefutable. It holds that you are the only conscious being in existence. The cosmos sprang into existence when you became sentient, and it will vanish when you die. As crazy as this proposition seems, it rests on a brute fact: each of us is sealed in an impermeable prison cell of subjective awareness. Even our most intimate exchanges might as well be carried out via Zoom.

You experience your own mind every waking second, but you can only infer the existence of other minds through indirect means. Other people seem to possess conscious perceptions, emotions, memories, intentions, just as you do, but you cannot be sure they do. You can guess how the world looks to me based on my behavior and utterances, including these words you are reading, but you have no firsthand access to my inner life. For all you know, I might be a mindless bot.

Natural selection instilled in us the capacity for a so-called theory of mind—a talent for intuiting others' emotions and intentions. But we have a countertendency to deceive one another and to fear we are being deceived. The ultimate deception would be pretending you are conscious when you are not.

The solipsism problem thwarts efforts to explain consciousness. Scientists and philosophers have proposed countless contradictory hypotheses about what consciousness is and how it arises. Panpsychists contend that all creatures and even inanimate matter—even a single proton!—possess consciousness. Hard-core materialists insist, conversely (and perversely), that not even humans are all that conscious.

The solipsism problem prevents us from verifying or falsifying these and other claims. I cannot be certain that you are conscious, let alone a jellyfish, bot or doorknob. As long as we lack what neuroscientist Christof Koch has called a consciousness meter—a device that can measure consciousness in the same way that a thermometer measures temperature—theories of consciousness will remain in the realm of pure speculation.

But the solipsism problem is far more than a technical philosophical matter. It is a paranoid but understandable response to the feelings of solitude that lurk within us all. Even if you reject solipsism as an intellectual position, you sense it, emotionally, whenever you feel estranged from others, whenever you confront the awful truth that you can never know—really know—another person, and no one can really know you.

Religion is one response to the solipsism problem. Our ancestors dreamed up a supernatural entity who bears witness to our innermost fears and desires. No matter how lonesome we feel, how alienated from our fellow humans, God is always there watching over us. He sees our souls, our most secret selves, and He loves us anyway. Wouldn't it be nice to think so?

The arts, too, can be seen as attempts to overcome the solipsism problem. The artist, musician, poet, novelist says, "This is how my life feels" or "This is how life might feel for another person." They help us imagine

Getty Images (*preceding pages*)

what it is like to be a Black woman trying to save her children from slavery or a Jewish ad salesman wandering through Dublin, wondering whether his wife is cheating on him. But to imagine is not to know.

Some of my favorite works of art dwell on the solipsism problem. In *I'm Thinking of Ending Things* and earlier films, as well as his novel *Antkind,* Charlie Kaufman depicts other people as projections of a disturbed protagonist. Kaufman no doubt hopes to help us, and himself, overcome the solipsism problem by venting his anxiety about it, but I find his dramatizations almost too evocative.

Love, ideally, gives us the illusion of transcending the solipsism problem. You feel you really know someone, from the inside out, and they know you. In moments of ecstatic sexual communion or mundane togetherness—while you're eating pizza and watching *The Alienist,* say—you fuse with your beloved. The barrier between you seems to vanish.

Inevitably, however, your lover disappoints, deceives, betrays you. Or, less dramatically, some subtle biocognitive shift occurs. You look at her as she nibbles her pizza and think, Who, what, is this odd creature? The solipsism problem has reemerged, more painful and suffocating than ever.

It gets worse. In addition to the problem of other minds, there is the problem of our own. As evolutionary psychologist Robert Trivers points out, we deceive ourselves at least as effectively as we deceive others. A corollary of this dark truth is that we know ourselves even less than we know others.

If a lion could talk, philosopher Ludwig Wittgenstein said, we couldn't understand it. The same is true, I suspect, of our own deepest selves. If you could eavesdrop on your subconscious, you would hear nothing but grunts, growls and moans—or perhaps the high-pitched squeaks of raw machine-code data coursing through a channel.

For the mentally ill, solipsism can become terrifyingly vivid. Victims of Capgras syndrome think that identical imposters have replaced their loved ones. If you have Cotard's delusion, also known as walking corpse syndrome, you become convinced that you are dead. A much more common disorder is derealization, which makes everything—you, others, reality as a whole—feel strange, phony, simulated.

Derealization plagued me throughout my youth. One episode was self-induced. Hanging out with friends in high school, I thought it would be fun to hyperventilate, hold my breath and let someone squeeze my chest until I passed out. When I woke up, I didn't recognize my buddies. They were demons jeering at me. For weeks after that horrifying sensation faded, everything still felt unreal, as if I were in a dreadful movie.

What if those afflicted with these alleged delusions actually see reality clearly? According to the Buddhist doctrine of anatta, the self does not really exist. When you try to pin down your own essence, to grasp it, it slips through your fingers.

We have devised methods for cultivating self-knowledge and quelling our anxieties, such as meditation and psychotherapy. But these practices strike me as forms of self-brainwashing. When we meditate or see a therapist, we are not solving the solipsism problem. We are merely training ourselves to ignore it, to suppress the horror and despair that it triggers.

NATURAL SELECTION INSTILLED IN US THE CAPACITY FOR A SO-CALLED THEORY OF MIND—A TALENT FOR INTUITING OTHERS' EMOTIONS AND INTENTIONS. BUT WE HAVE A COUNTERTENDENCY TO DECEIVE ONE ANOTHER AND TO FEAR WE ARE BEING DECEIVED.

We have also invented mythical places in which the solipsism problem vanishes. We transcend our solitude and merge with others into a unified whole. We call these places heaven, nirvana, the Singularity. But solipsism is a cave from which we cannot escape—except, perhaps, by pretending it does not exist. Or, paradoxically, by confronting it, the way Kaufman does. Knowing we are in the cave may be as close as we can get to escaping it.

Conceivably, technology could deliver us from the solipsism problem. Koch proposes that we all get brain implants with Wi-Fi so we can meld minds through a kind of high-tech telepathy. Philosopher Colin McGinn suggests a technique that involves "brain splicing," transferring bits of your brain into mine, and vice versa.

But do we really want to escape the prison of our subjective selves? The archnemesis of *Star Trek: The Next Generation* is the Borg, a legion of tech-enhanced humanoids who have fused into one big meta-entity. Borg members have lost their separation from one another and hence their individuality. When they meet ordinary humans, they mutter in a scary monotone, "You will be assimilated. Resistance is futile."

As hard as solitude can be for me to bear, I do not want to be assimilated. If solipsism haunts me, so does oneness, a unification so complete that it extinguishes my puny mortal self. Perhaps the best way to cope with the solipsism problem in this weird, lonely time is to imagine a world in which it has vanished. SA

John Horgan directs the Center for Science Writings at the Stevens Institute of Technology. His books include *The End of Science, The End of War* and *Mind-Body Problems,* available for free at mindbodyproblems.com. For many years he wrote the popular blog Cross Check for *Scientific American*.

How We Perceive the World

Our Inner Universes *page 38*

Why We Have Free Will *page 44*

The Brain Learns in Unexpected Ways *page 48*

The Brain's Social Road Maps *page 54*

Face Values *page 60*

Radical Change *page 68*

Illustration by Maria Corte

OUR INNER UNIVERSES

Reality is constructed by the brain, and no two brains are exactly alike

By Anil K. Seth

Illustration by Brook VanDevelder

"We do not see things as they are, we see them as we are."
—from *Seduction of the Minotaur,* by Anaïs Nin (1961)

ON THE 10TH OF APRIL 2019 POPE FRANCIS, PRESIDENT SALVA KIIR OF SOUTH SUDAN AND former rebel leader Riek Machar sat down together for dinner at the Vatican. They ate in silence, the start of a two-day retreat aimed at reconciliation from a civil war that had killed some 400,000 people since 2013. At about the same time in my laboratory at the University of Sussex in England, Ph.D. student Alberto Mariola was starting to work on an experiment in which volunteers experience being in a room they believe is there but is not. In psychiatry clinics across the globe, people arrive complaining that things no longer seem "real" to them, whether it is the world around them or their own selves. In the fractured societies in which we live, what is real—and what is not—seems to be increasingly up for grabs. Warring sides may experience and believe in different realities. Perhaps eating together in silence can help because it offers a small slice of reality that can be agreed on, a stable platform on which to build further understanding.

We need not look to war and psychosis to find radically different inner universes. In 2015 a badly exposed photograph of a dress tore across the Internet, dividing the world into those who saw it as blue and black (me included) and those who saw it as white and gold (half my lab). Those who saw it one way were so convinced they were right—that the dress truly was blue and black or white and gold—that they found it almost impossible to believe that others might perceive it differently.

We all know that our perceptual systems are easy to fool. The popularity of visual illusions is testament to this phenomenon. Things seem to be one way, and they are revealed to be another: two lines appear to be different lengths, but when measured they are exactly the same; we see movement in an image we know to be still. The story usually told about illusions is that they exploit quirks in the circuitry of perception, so that what we perceive deviates from what is there. Implicit in this story, however, is the assumption that a properly functioning perceptual system will render to our consciousness things precisely as they are.

The deeper truth is that perception is never a direct window onto an objective reality. All our perceptions are active constructions, brain-based best guesses at the nature of a world that is forever obscured behind a sensory veil. Visual illusions are fractures in the Matrix, fleeting glimpses into this deeper truth.

Take, for example, the experience of color—say, the bright red of the coffee mug on my desk. The mug really does seem to be red: its redness seems as real as its roundness and its solidity. These features of my experience seem to be truly existent properties of the world, detected by our senses and revealed to our mind through the complex mechanisms of perception.

Yet we have known since Isaac Newton that colors do not exist out there in the world. Instead they are cooked up by the brain from mixtures of different wavelengths of colorless electromagnetic radiation. Colors are a clever trick that evolution has hit on to help the brain keep track of surfaces under changing lighting conditions. And we humans can sense only a tiny slice of the full electromagnetic spectrum, nestled between the lows of infrared and the highs of ultraviolet. Every color we perceive, every part of the totality of each of our visual worlds, comes from this thin slice of reality.

Just knowing this is enough to tell us that perceptual experience cannot be a comprehensive representation of an external objective world. It is both less than that and more than that. The reality we experience—the way things *seem*—is not a direct reflection of what is actually out there. It is a clever construction by the brain, for the brain. And if my brain is different from your brain, my reality may be different from yours, too.

THE PREDICTIVE BRAIN

IN PLATO'S *Allegory of the Cave,* prisoners are chained to a blank wall all their lives, so that they see only the play of shadows cast by objects passing by a fire behind them, and they give the shadows names because for them the shadows are what is real. A thousand years later, but still a thousand years ago, Arabian scholar Ibn al-Haytham wrote that perception, in the here and now, depends on processes of "judgment and inference" rather than involving direct access to an objective reality. Hundreds of years later again Immanuel Kant realized that the chaos of unrestricted sensory data would always remain meaningless without being given structure by preexisting conceptions or "beliefs," which for him included a priori frameworks such as space and time. Kant's term "noumenon" refers to a "thing in itself"—*Ding an sich*—an objective reality that will always be inaccessible to human perception.

Today these ideas have gained a new momentum through an influential collection of theories that turn on the idea that the brain is a kind of prediction machine and that perception of the world—and of the self within it—is a process of brain-based prediction about the causes of sensory signals.

These new theories are usually traced to German physicist and physiologist Hermann von Helmholtz, who in the late 19th century proposed that perception is a process of unconscious inference. Toward the end of the 20th century Helmholtz's notion was taken up by cognitive scientists and artificial-intelligence researchers, who reformulated it in terms of what is now generally known as predictive coding or predictive processing.

The central idea of predictive perception is that the brain is attempting to figure out what is out there in the world (or in here, in the body) by continually making and updating best guesses about the causes of its sensory inputs. It forms these best guesses by combining prior expectations or "beliefs" about the world, together with incoming sensory data, in a way that takes into account how reliable the sensory signals are. Scientists usually conceive of this process as a form of Bayesian inference, a framework that specifies how to update beliefs or best guesses with new data when both are laden with uncertainty.

In theories of predictive perception, the brain approximates this kind of Bayesian inference by continually generating predictions about sensory signals and comparing these predictions with

the sensory signals that arrive at the eyes and the ears (and the nose and the fingertips, and all the other sensory surfaces on the outside and inside of the body). The differences between predicted and actual sensory signals give rise to so-called prediction errors, which are used by the brain to update its predictions, readying it for the next round of sensory inputs. By striving to minimize sensory-prediction errors everywhere and all the time, the brain implements approximate Bayesian inference, and the resulting Bayesian best guess is what we perceive.

To understand how dramatically this perspective shifts our intuitions about the neurological basis of perception, it is helpful to think in terms of bottom-up and top-down directions of signal flow in the brain. If we assume that perception is a direct window onto an external reality, then it is natural to think that the content of perception is carried by bottom-up signals—those that flow from the sensory surfaces inward. Top-down signals might contextualize or finesse what is perceived, but nothing more. Call this the "how things seem" view because it seems as if the world is revealing itself to us directly through our senses.

POORLY EXPOSED photograph of a dress appears blue and black to some people, white and gold to others.

The prediction machine scenario is very different. Here the heavy lifting of perception is performed by the top-down signals that convey perceptual predictions, with the bottom-up sensory flow serving only to calibrate these predictions, keeping them yoked, in some appropriate way, to their causes in the world. In this view, our perceptions come from the inside out just as much as, if not more than, from the outside in. Rather than being a passive registration of an external objective reality, perception emerges as a process of active construction—a controlled hallucination, as it has come to be known.

Why controlled hallucination? People tend to think of hallucination as a kind of false perception, in clear contrast to veridical, true-to-reality, normal perception. The prediction machine view suggests instead a continuity between hallucination and normal perception. Both depend on an interaction between top-down, brain-based predictions and bottom-up sensory data, but during hallucinations, sensory signals no longer keep these top-down predictions appropriately tied to their causes in the world. What we call hallucination, then, is just a form of uncontrolled perception, just as normal perception is a controlled form of hallucination.

This view of perception does not mean that nothing is real. Writing in the 17th century, English philosopher John Locke made an influential distinction between "primary" and "secondary" qualities. Primary qualities of an object, such as solidity and occupancy of space, exist independently of a perceiver. Secondary qualities, in contrast, exist only in relation to a perceiver—color is a good example. This distinction explains why conceiving of perception as controlled hallucination does not mean it is okay to jump in front of a bus. This bus has primary qualities of solidity and space occupancy that exist independently of our perceptual machinery and that can do us injury. It is the way in which the bus appears to us that is a controlled hallucination, not the bus itself.

TRIPPING IN THE LAB

A GROWING BODY of evidence supports the idea that perception is controlled hallucination, at least in its broad outlines. A 2015 study by Christoph Teufel of Cardiff University in Wales and his colleagues offers a striking example. In this study, the ability to recognize so-called two-tone images was evaluated in patients with early-stage psychosis who were prone to hallucinations.

Take a look at the top photograph on page 43—a sample of a two-tone image. Probably all you will see is a bunch of black-and-white splotches. Now look at the image at the bottom of that page. Then have another look at the first photo; it ought to look rather different. Where previously there was a splotchy mess, there are now distinct objects, and something is happening.

What I find remarkable about this exercise is that in your second examination of the top image, the sensory signals arriving at your eyes have not changed at all from the first time you saw it. All that has changed are your brain's predictions about the causes of these sensory signals. You have acquired a new high-level perceptual expectation, and this changes what you consciously see.

If you show people many two-tone images, each followed by the full picture, they might subsequently be able to identify a good proportion of two-tone images, though not all of them. In Teufel's study, people with early-stage psychosis were better at recognizing two-tone images after having seen the full image than were healthy control subjects. In other words, being hallucination-prone went along with perceptual priors having a stronger effect on perception. This is exactly what would be expected if hallucinations in psychosis depended on an overweighting of perceptual priors so that they overwhelmed sensory prediction errors, unmooring perceptual best guesses from their causes in the world.

Recent research has revealed more of this story. In a 2021 study, Biyu He of New York University and her colleagues had neurosurgical patients look at ambiguous images, such as a Necker cube, that constantly flip between two different appearances even though the sensory input remains the same. By analyzing the signals recorded from within the patients' brains, they discovered that information flowed more strongly in a top-down, "inside-out" direction when the perceived appearance was consistent with the patients' biases, as would be expected if perceptual predictions were strong in this case. And when the perceived appearance was inconsistent with preexisting biases, information flow was stronger in the bottom-up direction, suggest-

Swiked.tumblr.com

ing a "prediction error" signal. This is an exciting new development in mapping the brain basis of controlled hallucinations.

In my lab we have taken a different approach to exploring the nature of perception and hallucination. Rather than looking into the brain directly, we decided to simulate the influence of overactive perceptual priors using a unique virtual-reality setup masterminded by our resident VR guru, Keisuke Suzuki. We call it, with tongue firmly in cheek, the "hallucination machine."

Using a 360-degree camera, we first recorded panoramic video footage of a busy square in the University of Sussex campus on a Tuesday at lunchtime. We then processed the footage through an algorithm based on Google's AI program DeepDream to generate a simulated hallucination. What happens is that the algorithm takes a so-called neural network—one of the workhorses of AI—and runs it backward. The network we used had been trained to recognize objects in images, so if you run it backward, updating the network's input instead of its output, the network effectively projects what it "thinks" is there onto and into the image. Its predictions overwhelm the sensory inputs, tipping the balance of perceptual best guessing toward these predictions. Our particular network was good at classifying different breeds of dogs, so the video became unusually suffused by dog presences.

OUR PERCEPTIONS COME FROM THE INSIDE OUT JUST AS MUCH AS, IF NOT MORE THAN, FROM THE OUTSIDE IN.

Many people who have viewed the processed footage through the VR headset have commented that the experience is rather reminiscent not of the hallucinations of psychosis but of the exuberant phenomenology of psychedelic trips.

More recently, we have been implementing the hallucination machine in different ways to simulate different kinds of altered visual experience. By extending our algorithm to include two coupled neural networks—a "discriminator network" much like the one in our original study and a "generator" network that has been trained to reproduce ("generate") its input image—we have been able to model different types of hallucination. For example, we have modeled the complex hallucinatory experiences reported by people with Parkinson's disease and some forms of dementia; the patterned, geometric hallucinations that occur after the loss of foveal vision, as happens in Charles Bonnet syndrome; and a range of psychedeliclike hallucinations. We hope that by understanding hallucinations better, we will be able to understand normal experience better, too, because predictive perception is at the root of all our perceptual experience.

THE PERCEPTION OF REALITY

ALTHOUGH THE HALLUCINATION machine is undoubtedly trippy, people who experience it are fully aware that what they are experiencing is not real. Indeed, despite rapid advances in VR technology and computer graphics, no current VR setup delivers an experience that is sufficiently convincing to be indistinguishable from reality.

This is the challenge we took up when designing a new "substitutional reality" setup at Sussex—the one we were working on when Pope Francis convened the retreat with Salva Kiir and Riek Machar. Our aim was to create a system in which volunteers would experience an environment as being real—and believe it to be real—when in fact it was not real.

The basic idea is simple. We again prerecorded some panoramic video footage, this time of the interior of our VR lab rather than of an outside campus scene. People coming to the lab are invited to sit on a stool in the middle of the room and to put on a VR headset that has a camera attached to the front. They are encouraged to look around the room and to see the room as it actually is, via the camera. But at some point, without telling them, we switch the feed so that the headset now displays not the live real-world scene but rather the prerecorded panoramic video. Most people in this situation continue to experience what they are seeing as real even though it is now a fake prerecording. (This is actually very tricky to pull off in practice—it requires careful color balancing and alignment to avoid people noticing any difference that would tip them off to the shift.)

I find this result fascinating because it shows that it is possible to have people experience an unreal environment as being fully real. This demonstration alone opens new frontiers for VR research: we can test the limits of what people will experience, and believe, to be real. It also allows us to investigate how experiencing things as being real can affect other aspects of perception. Right now we are running an experiment to find out whether people are worse at detecting unexpected changes in the room when they believe that what they are experiencing is real. If things do turn out this way (the study is still ongoing, despite being heavily delayed by a global pandemic), that finding would support the idea that the perception of things as being real itself acts as a high-level prior that can substantively shape our perceptual best guesses, affecting the contents of what we perceive.

THE REALITY OF REALITY

THE IDEA THAT THE WORLD of our experience might not be real is an enduring trope of philosophy and science fiction, as well as of late-night pub discussions. Neo in *The Matrix* takes the red pill, and Morpheus shows him how what he thought was real is an elaborate simulation, while the real Neo lies prone in a human body farm, a brain-in-a-vat power source for a dystopian AI. Philosopher Nick Bostrom of the University of Oxford has famously argued, based largely on statistics, that we are likely to be living inside a computer simulation created in a posthuman age. I disagree with this argument in part because it assumes that consciousness can be simulated—I do not think that this is a safe assumption—but it is thought-provoking nonetheless.

Although these chunky metaphysical topics are fun to chew on, they are probably impossible to resolve. Instead what we have been exploring throughout this article is the relation between appearance and reality in our conscious perceptions, where part of this appearance is the appearance of being real itself.

The central idea here is that perception is a process of active interpretation geared toward adaptive interaction with the world through the body rather than a re-creation of the world within the mind. The contents of our perceptual worlds are controlled hallucinations, brain-based best guesses about the ultimately unknowable causes of sensory signals. For most of us, most of the time, these controlled hallucinations are experienced as real. As Canadian rapper and science communicator Baba Brinkman suggested to me,

when we agree about our hallucinations, maybe that is what we call reality.

But we do not always agree, and we do not always experience things as real. People with dissociative psychiatric conditions such as derealization or depersonalization syndrome report that their perceptual worlds, even their own selves, lack a sense of reality. Some kinds of hallucination, various psychedelic hallucinations among them, combine a sense of unreality with perceptual vividness, as does lucid dreaming. People with synesthesia consistently have additional sensory experiences, such as perceiving colors when viewing black letters, which they recognize as not real. Even with normal perception, if you look directly at the sun you will experience the subsequent retinal afterimage as not being real. There are many such ways in which we experience our perceptions as not fully real.

TWO-TONE IMAGE looks like a mess of black-and-white splotches, until you see the full image (*below*).

What this means to me is that the property of realness that attends most of our perceptions should not be taken for granted. It is another aspect of the way our brain settles on its Bayesian best guesses about its sensory causes. One might thus ask what purpose it serves. Perhaps the answer is that a perceptual best guess that includes the property of being real is usually more fit for purpose—that is, better able to guide behavior—than one that does not. We will behave more appropriately with respect to a coffee cup, an approaching bus or our partner's mental state when we experience it as really existing.

But there is a trade-off. As illustrated by the dress illusion, when we experience things as being real, we are less able to appreciate that our perceptual worlds may differ from those of others. (A popular explanation for the differing perceptions of the garment holds that people who spend most of their waking hours in daylight see it as white and gold; night owls, who are mainly exposed to artificial light, see it as blue and black.) And even if these differences start out small, they can become entrenched and reinforced as we proceed to harvest information differently, selecting sensory data that are best aligned with our individual emerging models of the world and then updating our perceptual models based on these biased data. We are all familiar with this process from the echo chambers of social media and the newspapers we choose to read. I am suggesting that the same principles apply also at a deeper level, underneath our sociopolitical beliefs, right down to the fabric of our perceptual realities. They may even apply to our perception of being a self—the experience of being me or of being you—because the experience of being a self is itself a perception.

This is why understanding the constructive, creative mechanisms of perception has unexpected social relevance. Perhaps once we can appreciate the diversity of experienced realities scattered among the billions of perceiving brains on this planet, we will find new platforms on which to build a shared understanding and a better future—whether between sides in a civil war, followers of different political parties, or two people sharing a house and faced with washing the dishes. SA

Anil K. Seth is a professor of cognitive and computational neuroscience at the University of Sussex in England. His research focuses on the biological basis of consciousness, and he is author of *Being You—A New Science of Consciousness* (Dutton, 2021).

PERCEPTUAL SHIFT: Viewing this photograph changes what one consciously sees in the two-tone image (*above*).

Richard Armstrong/EyeEm/Getty Images

why we have free will

Neurons fire in your head before you become aware that you have made a decision. But this discovery does not mean you are a "biochemical puppet"

By Eddy Nahmias

Illustration by Simon Prades

ONE FALL NIGHT I LAY AWAKE WONDERING HOW I SHOULD begin this essay. I imagined a variety of ways I could write the first sentence and the next and the one after that. Then I thought about how I could tie those sentences to the following paragraph and the rest of the article. The pros and cons of each of those options circled back and forth in my head, keeping me from drifting off to sleep. As this was happening, neurons were buzzing away in my brain. Indeed, that neural activity explains why I imagined these options, and it explains why I am writing these very words. It also explains why I have free will.

Increasingly, neuroscientists, psychologists and pundits say that I am wrong. Invoking a number of widely cited neuroscientific studies, they claim that unconscious processes drove me to select the words I ultimately wrote. Their arguments suggest our conscious deliberation and decisions happen only after neural gears below the level of our conscious awareness have already determined what we will choose. And they conclude that because "our brains make us do it"—choosing for us one option over another—free will is nothing more than an illusion.

The experiments most often cited to show that our brains take charge behind the scenes were carried out by the late Benjamin Libet in the 1980s at the University of California, San Francisco. There he instructed study participants outfitted with electrodes on their heads to flick their wrists whenever they felt like it. The electrodes detected fluctuations in electrical activity called readiness potentials that occurred about half a second before people made the flicking motion. But participants became aware of their intentions to move only about a quarter of a second before the movement, leading to the conclusion that their brains had decided before they became aware of what had happened. In essence, unconscious brain processes were in the driver's seat.

More recent studies using functional MRI have suggested the unconscious roots of our decisions begin even earlier. In research published in 2013, neuroscientist John-Dylan Haynes of the Bernstein Center for Computational Neuroscience Berlin and his colleagues had volunteers decide whether to add or subtract two numbers while in the fMRI scanner. They found patterns of neural activity that were predictive of whether subjects would choose to add or subtract that occurred four seconds before those subjects were aware of making the choice.

These studies—and others like them—have led to sweeping pronouncements that free will is dead. "Our decisions are predetermined unconsciously a long time before our consciousness kicks in," Haynes commented to *New Scientist,* adding that "it seems that the brain is making the decision before the person." Others share his opinion. Evolutionary biologist Jerry Coyne has written: "So it is with all of our ... choices: not one of them results from a free and conscious decision on our part. There is no freedom of choice, no free will." Neuroscientist Sam Harris has concluded from these findings that we are "biochemical puppets": "If we were to detect [people's] conscious choices on a brain scanner seconds before they were aware of them … this would directly challenge their status as conscious agents in control of their inner lives."

But does the research really show that all our conscious deliberation and planning is just a by-product of unconscious brain activity, having no effect on what we do later on? No, it does not. There are several reasons to think that those who insist that free will is a mirage are misguided.

NOT SO FAST

I CALL THOSE who contend that science shows that free will is an illusion "willusionists." There are many reasons to be wary of the willusionists' arguments. First, neuroscience currently lacks the technical sophistication to determine whether neural activity underlying our imagining and evaluating of future options has any impact on which option we then carry out minutes, hours or days later. Instead the research discussed by willusionists fails to clearly define the border between conscious and unconscious actions.

Consider the Libet experiment. It began with study participants preparing consciously to make a series of repetitive and unplanned actions. When the experiment began, they flexed their wrists when a desire arose spontaneously. The neural activity involved in the conscious planning presumably influenced the later unconscious initiation of movements, revealing an interaction between conscious and unconscious brain activity.

Similarly, the 2011 Haynes study, in which people randomly picked whether to add or subtract over the course of many trials, fails to provide convincing evidence against free will. Early brain activity that occurred four seconds before participants were aware of making a choice may be an indication of unconscious biases toward one choice or the other.

But this early brain activity predicted a choice with an accuracy only 10 percent better than random chance. Brain activity cannot, in general, *settle* our choices four seconds before we act, because we can react to changes in our situation in less time than that. If we could not, we would all have died in car crashes by now! Unconscious neural activity, however, can prepare us to take an action by cuing us to consciously monitor our actions to let us adjust our behavior as it occurs.

Willusionists also point to psychological research showing that we have less conscious control over our actions than we think. It is true that we are often influenced unknowingly by subtle features of our environment and by emotional or cognitive biases. Until we understand them, we are not free to try to counteract them. This is one reason I think we have *less* free will than many people tend to believe. But there is a big difference between less and none at all.

The Libet and Haynes research deals with choices that people make without conscious deliberation at the time of action. Everyone performs repetitive or habitual behaviors, sometimes quite sophisticated ones that do not require much thought because the behaviors have been learned. You put your key in the lock. A shortstop dives for a ground ball. A pianist becomes immersed in playing Beethoven's *Moonlight Sonata*.

The reflexive turning of the key, the lunging for the ball, or the depressing of the white and black keys requires a particular type of mental processing. What I was doing on that sleepless night—conscious consideration of alternative options—is a wholly different activity from engaging in practiced routines. A body of psychological research shows that conscious, purposeful processing of our

thoughts likely does make a difference in what we do.

This work indicates that intentions we formulate to carry out specific tasks in particular circumstances—what psychologists call "implementation intentions"—increase the likelihood that we will complete the planned behavior. A study performed by psychologist Peter Gollwitzer of New York University and his colleagues revealed that dieters who consciously formed an intention to ignore thoughts about tempting foods whenever they came to mind then ate less of those foods than those dieters who simply set the goal to lose weight.

Psychologist Roy F. Baumeister, then at Florida State University, and his colleagues demonstrated that conscious reasoning improves performance on logical and linguistic tasks and that it helps in learning from past mistakes and overriding impulsive behaviors. In addition, the late Walter Mischel of Columbia University found that our ability to willfully distract ourselves from a temptation is crucial for self-control.

Every one of us carries out actions every day that we have consciously planned for ourselves. It is possible that the neural activity that carries out this planning has no effect on what we do or that it just concocts stories after the fact to explain to ourselves and others what we did. But that would make little evolutionary sense. The brain makes up only 2 percent of the human body's weight but consumes 20 percent of its energy. There would be strong evolutionary pressure against neural processes that enable intricate conscious thought yet are irrelevant to our behavior. The brain circuits responsible for my imagining that this is the best way to write this essay are likely causing it to turn out this way.

FREE WILL IN THE BRAIN?

WILLUSIONISTS, however, suggest brain processing responsible for conscious thinking simply cannot count as free will. They often say that people who believe in free will must be "dualists" who are convinced that the mind somehow exists as a nonphysical entity, separate from the brain. "Free will is the idea that we make choices and have thoughts independent of anything remotely resembling a physical process," wrote neuroscientist Read Montague in 2008. And Coyne has claimed that "true 'free will' … would require us to step outside of our brain's structure and modify how it works."

It is true that some people think of free will in this way. But there is no good reason to do so. Most philosophical theories develop a view of free will that is consistent with a scientific understanding of human nature. And despite willusionists' claims, studies suggest most people accept that we can have free will even if our mental activity is carried out entirely by brain activity. If most people are not committed to a dualist view about free will, then it is a mistake to tell them that free will is an illusion based on the scientific view that dualism is false.

One way to test people's assumptions about free will is to describe the possibility of brain-imaging technology that would allow perfect prediction of actions based on information about prior brain activity. In fact, Harris has suggested this scenario "would expose this feeling [of free will] for what it is: an *illusion*."

To see whether people's belief in free will would be challenged by the knowledge that the brain is engaged in unconscious information processing that predicts behavior, Jason Shepard of Life University in Georgia, Shane Reuter, then at Washington University in St. Louis, and I performed a series of experiments in which we presented people with detailed scenarios describing futuristic brain-imaging technology, as posited by Harris.

Hundreds of students at Georgia State University participated in the studies. They read about a woman named Jill who, in the distant future, wore a brain-imaging cap for a month. Using information from the brain scanner, neuroscientists predicted everything she thought and did, even when she tried to fool the system. The scenario concluded that "these experiments confirm that all human mental activity just *is* brain activity such that everything that any human thinks or does could be predicted ahead of time based on their earlier brain activity."

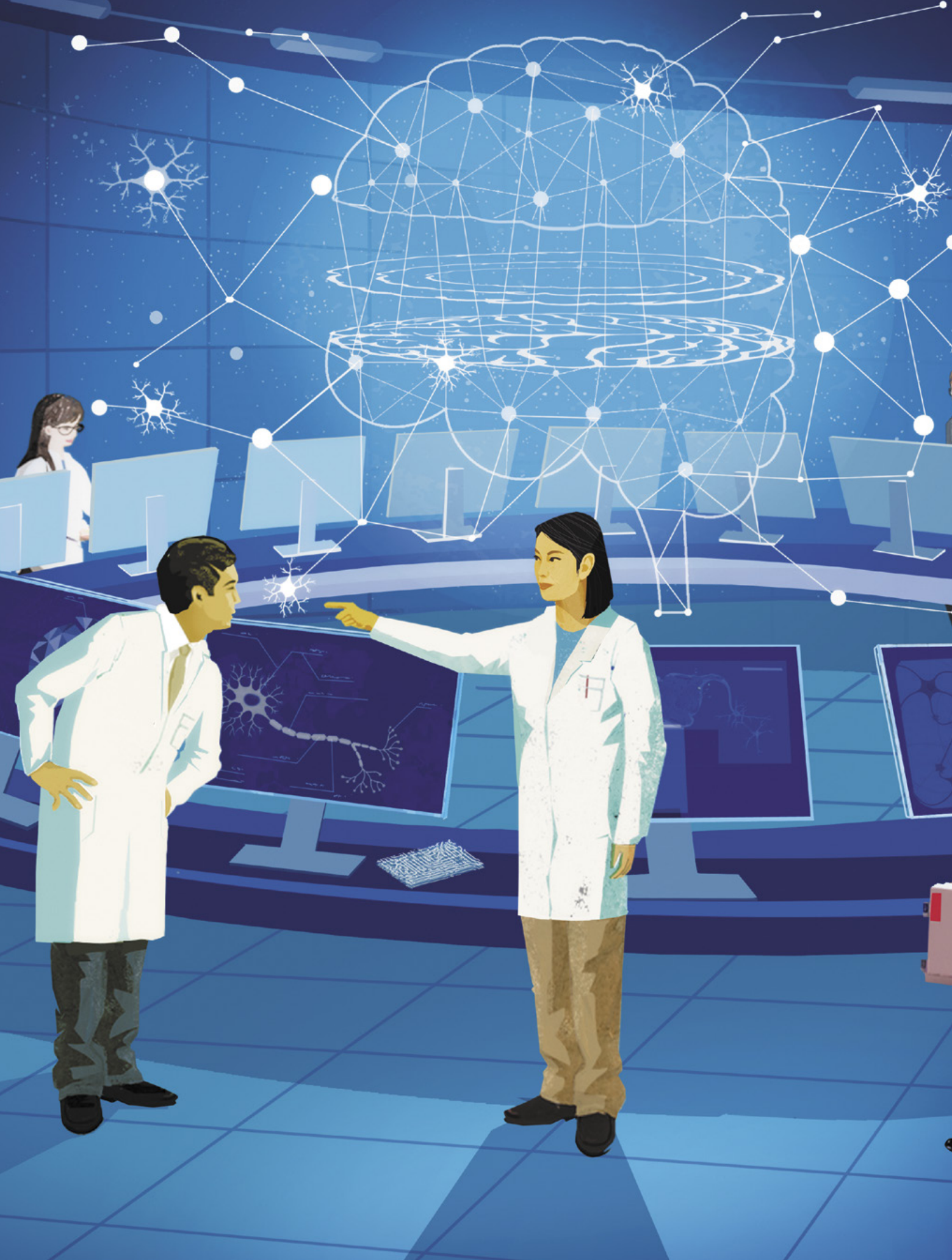
More than 80 percent of the participants reported that they believed that such future technology was possible, yet 87 percent of them responded that Jill still had free will. They were also asked whether the existence of such technology would indicate that individuals lack free will. Roughly 75 percent disagreed. Further results showed that a significant majority felt that as long as the technology did not allow people's brains to be manipulated such that their decisions could be controlled by others, they would have free will and be morally responsible for their behavior.

Most participants in the experiments seem to think that the hypothetical brain scanner is just recording the brain activity that is Jill's conscious reasoning and consideration about what to decide. Rather than taking this to mean that Jill's brain is making her do something—and that she has no free will—they may just be thinking that the brain scanner is simply detecting how free will works in the brain.

Why, then, do willusionists believe the opposite? It may have to do with the current state of knowledge. Until neuroscience is able to explain consciousness—which will require a theory to explain how our mind is neither reducible to nor distinct from the workings of our brain—it is tempting to think, as the willusionists seem to, that if the brain does it all, there is nothing left for the conscious mind to do.

As neuroscience advances and imaging technology improves, these developments should help reveal more precisely how much conscious control we have and to what extent our actions are governed by processes beyond our control. Finding resolutions for these questions about free will is important. Our legal system—and the moral basis for many of our society's institutions—requires a better understanding of when people are, and are not, responsible for what they do. SA

Eddy Nahmias is a professor in the department of philosophy and the Neuroscience Institute at Georgia State University.



The Brain Learns in Unexpected Ways

Neuroscientists have discovered a set of unfamiliar cellular mechanisms for making fresh memories

By R. Douglas Fields

Illustration by Eva Vazquez

Our concepts of how the two and a half pounds of flabby flesh between our ears accomplish learning date to Ivan Pavlov's classic experiments, where he found that dogs could learn to salivate at the sound of a bell. In 1949 psychologist Donald Hebb adapted Pavlov's "associative learning rule" to explain how brain cells might acquire knowledge. Hebb proposed that when two neurons fire together, sending off impulses simultaneously, the connections between them—the synapses—grow stronger. When this happens, learning has taken place. In the dogs' case, it would mean the brain now knows that the sound of a bell is followed immediately by the presence of food. This idea gave rise to an oft-quoted axiom: "Synapses that fire together wire together."

The theory proved sound, and the molecular details of how synapses change during learning have been described in depth. But not everything we remember results from reward or punishment, and in fact, most experiences are forgotten. Even when synapses do fire together, they sometimes do not wire together. What we retain depends on our emotional response to an experience, how novel it is, where and when the event occurred and our level of attention and motivation during the event, and we

process these thoughts and feelings while asleep. A narrow focus on the synapse has given us a mere stick-figure conception of how learning and the memories it engenders work.

It turns out that strengthening a synapse cannot produce a memory on its own, except for the most elementary reflexes in simple circuits. Vast changes throughout the expanse of the brain are necessary to create a coherent memory. Whether you are recalling last night's conversation with dinner guests or using an acquired skill such as riding a bike, the activity of millions of neurons in many different regions of your brain must become linked to produce a coherent memory that interweaves emotions, sights, sounds, smells, event sequences and other stored experiences. Because learning encompasses so many elements of our experiences, it must incorporate different cellular mechanisms beyond the changes that occur in synapses. This recognition has led to a search for new ways to understand how information is transmitted, processed and stored in the brain to bring about learning. In the past 10 years neuroscientists have come to realize that the iconic "gray matter" that makes up the brain's outer surface—familiar from graphic illustrations found everywhere from textbooks to children's cartoons—is not the only part of the organ involved in the inscription of a permanent record of facts and events for later recall and replay. It turns out that areas below the deeply folded, gray-colored surface also play a pivotal role in learning. In just the past few years a series of studies from my laboratory and others has elucidated these processes, which could point to new ways of treating psychiatric and developmental disorders that occur when learning impairments arise.

If synaptic changes alone do not suffice, what does happen inside your brain when you learn something new? Magnetic resonance imaging methods now enable researchers to see through a person's skull and examine the brain's structure. In scrutinizing MRI scans, investigators began to notice differences in the brain structure of individuals with specific highly developed skills. Musicians, for example, have thicker regions of auditory cortex than nonmusicians. At first, researchers presumed that these subtle differences must have predisposed clarinetists and pianists to excel at their given skills. But subsequent research found that learning changes the structure of the brain.

The kind of learning that leads to alterations in brain tissue is not limited to repetitive sensorimotor skills such as playing a musical instrument. Neuroscientist Bogdan Draganski, currently at the University of Lausanne in Switzerland, and his colleagues witnessed increases in the volume of gray matter in medical students' brains after they studied for an examination. Many different cellular changes could expand gray matter volume, including the birth of new neurons and of nonneuronal cells called glia. Vascular changes and the sprouting and pruning of axons and dendrites that extend from the main body of a neuron could also do the same. Remarkably, physical changes in the brain can happen much faster during learning than might be expected. Yaniv Assaf of Tel Aviv University and his colleagues showed that 16 laps around a race track in a computerized video game were enough to cause changes in new players' hippocampal brain region. Structural alterations in the hippocampus in these gamers make sense because this brain region is critical for spatial learning for navigation. In other studies, Assaf and, separately, Heidi Johansen-Berg of the University of Oxford were surprised to find changes in unexpected parts of the brain, including regions that have no neurons or synapses—areas known as white matter.

DEEP LEARNING

CONSCIOUSNESS ARISES from the cerebral cortex, the three-millimeter-thick outer layer of the human brain, so this gray matter layer is where most researchers expected to find learning-induced modifications. But below the surface layer, billions of tightly packed bundles of axons (nerve fibers), much like tightly wound fibers under the leather skin of a baseball, connect neurons in the gray matter into circuits.

These fiber bundles are white because the axons are coated with a fatty substance called myelin, which acts as electrical insulation and boosts the speed of transmission by 50 to 100 times. White matter injury and disease are important areas of research, but until recently little attention had been given in these investigations to a possible role of myelin in information processing and learning.

In the past 10 or so years studies have begun to find differences in white matter in brain scans of experts with a variety of skills, including people with high proficiency in reading and arithmetic. Expert golfers and trained jugglers also show differences in white matter compared with novices, and white matter volume has even been associated with IQ. If information processing and learning arise from the strengthening of synaptic connections between neurons in gray matter, why does learning affect the brain's subsurface cabling?

A possible answer began to emerge from cellular studies in my lab investigating how synapses—but also other brain areas—change during learning. The reason for looking beyond the synapse was that most of the drugs we have for treating neurological and psychological disorders work by altering synaptic transmission, and there is a pressing need for more effective agents. The present focus on synaptic transmission might cost us opportunities for better treatments for dementia, depression, schizophrenia or post-traumatic stress disorder (PTSD).

In the early 1990s my lab at the National Institutes of Health and others began to explore the possibility that glia might be able to sense information flowing through neural networks and alter it to improve performance. Experimental evidence that has accumulated since then shows that all types of glial cells respond to neural activity and can modify information transmission in the brain. One of the most surprising of these new findings involves myelin.

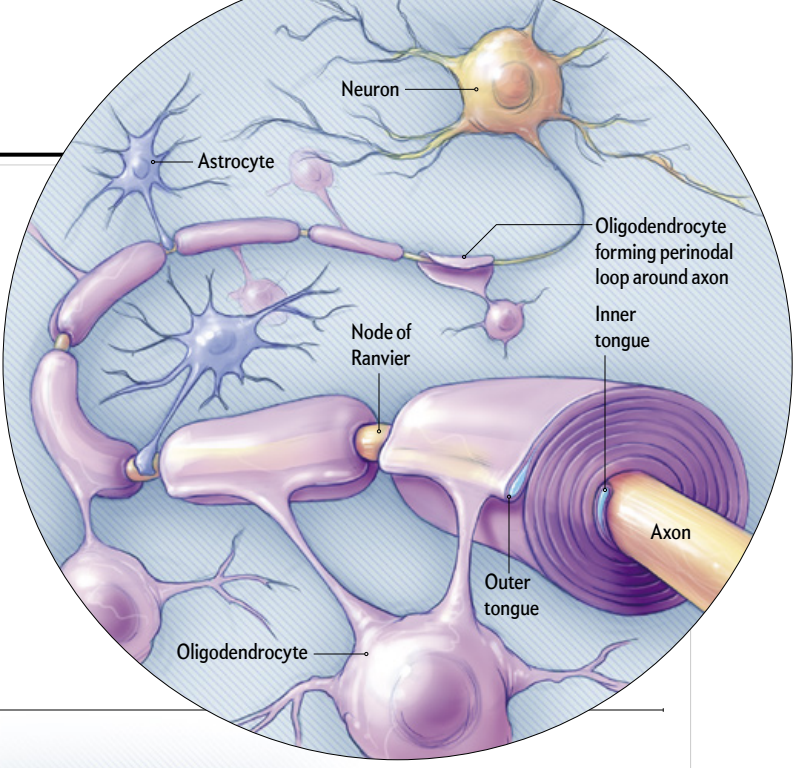
Myelin insulation is formed by layers of cell membrane wrapped around axons like electrical tape. In the brain and spinal cord, octopus-shaped glial cells (oligodendrocytes) do the wrapping. In the limbs and trunk, sausage-shaped glial cells (Schwann cells) perform the same task. Many oligodendrocytes grip an axon and wrap layers of myelin around it in segments, like the stacked hands of baseball players gripping a bat to determine which team bats first. The tiny gap between two myelin segments exposes a one-micron section of bare axon where ion channels that generate electrical impulses become concentrated. These spaces, known as the nodes of Ranvier, act like bioelectric repeaters to relay an electrical impulse from node to node down the axon. The speed of impulse transmission increases as more layers of myelin are wrapped around the axon, protecting it better against voltage loss. Also, as

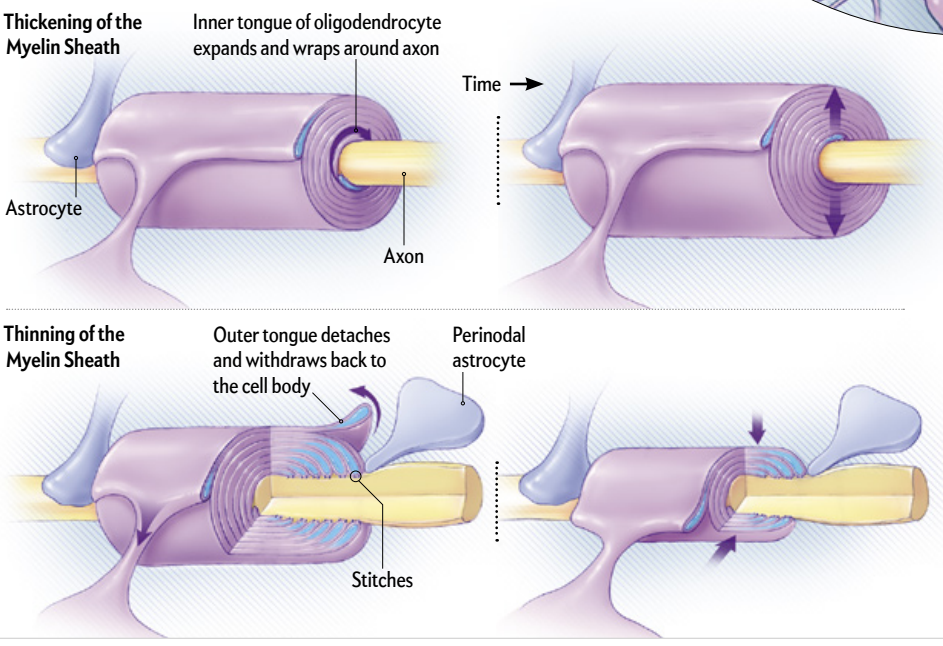
Axon Insulation

Neuroscience textbooks recount that the connecting points between neurons—the synapses—undergo alterations when learning takes place. But new research shows that changes also occur in myelin, part of the white matter that forms a sheath around the long filaments (axons) that stretch out from the cell body of a neuron.

Worker Cells

Insulating sheaths made of fatty, white myelin control the rate at which electrical signals travel along axons. Cells called oligodendrocytes loop around and wrap myelin on an axon—and, in some cases, remove it. Small gaps in myelin (nodes of Ranvier) contain ion channels that generate electrical impulses. Another cell type, the perinodal astrocyte, stops the secretion of the myelin-removing thrombin (*not shown*).





Wrapping and Unwrapping a Neuron

Oligodendrocytes start wrapping myelin around axons in electrically active neurons. The degree of myelination controls how fast a signal travels along an axon, with thicker sheaths producing speedier transmissions. The enzyme thrombin cuts the stitches that bind myelin to the axon, and the perinodal astrocyte brings this process to a halt to procure the desired thickness. Varying myelin's depth ensures that dispersed signals arrive at a neural relay point at the same time, enhancing performance on a new task.

a node of Ranvier becomes squeezed more tightly by the adjoining myelin segments, an electrical impulse is initiated more rapidly because it takes less time to charge the smaller amount of nodal membrane to the voltage that triggers ion channels to open and generate an impulse.

Disorders that damage myelin, such as multiple sclerosis and Guillain-Barré syndrome, can cause serious disability because neural impulse transmission fails when the insulation is damaged. But until recently, the idea that myelin might be modified routinely by neural impulses was not widely accepted. And even if myelin structure changed, how and why would this improve performance and learning?

The explanation was hiding in plain sight. It loops back to the old maxim about neurons firing and wiring together. In any complex information or transportation network, the time of arrival at network relay points is critical—think of missing a connection because your flight arrives too late.

How, then, does the transmission speed in every link in the human brain get timed appropriately so that an impulse arrives just when needed? We know that electrical signals shuffle along at the pace of a slow walk in some axons but blaze away at the speed of a race car in others. Signals from two axons that converge on neurons that act as relay points will not arrive together unless the travel time from their input source is optimized to compensate for differences in the lengths of the two axons and the speed at which impulses travel along each link.

Because myelin is the most effective means of speeding impulse transmission, axon myelination promotes optimal information transmission through a network. If oligodendrocytes sense and respond to the information traffic flowing through neural circuits, then myelin formation and the way it adjusts impulse-transmission speed could be controlled by feedback from the axon. But how can myelinating glia detect neural impulses flowing through axons?

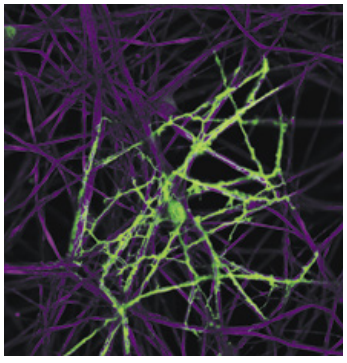
Source: "Treadmilling Model for Plasticity of the Myelin Sheath," by R. Douglas Fields and Dipankar J. Dutta, in *Trends in Neurosciences*, Vol. 42, No. 7; July 2019

Illustration by David Cheney

SIGNAL TRANSMISSION

OVER THE PAST TWO DECADES our research and that of other labs has succeeded in identifying many neurotransmitters and other signaling molecules that convey to glia the presence of electrical activity in the axon to stimulate myelination. Our experiments have shown that when a neuron fires, neurotransmitters are released not only at synapses but also all along the axon. We found that the "tentacles" of the octopuslike oligodendrocytes probe bare sections of axons in search of neurotransmitters being released from axons firing. When a single tentacle touches an axon that is firing, it forms a "spot weld" contact, which enables communication between the axon and the oligodendrocyte. The oligodendrocyte begins to synthesize myelin at that spot and wrap it around the axon.

When we gave oligodendrocytes in cell culture the choice of myelinating electrically active axons or ones treated with botulinum toxin to prevent the release of neurotransmitters, the oligodendrocytes opted for the electrically active axons over the silent ones by a factor of eight to one. So it may be that as a person learns to play "Für Elise" on the piano, bare axons are wrapped with myelin or the volume of existing sheaths is increased in circuits that are activated repetitively during practice, which speeds information flow through brain networks. New myelin then shows up on an MRI as changes in white matter tracts in parts of the brain that are necessary for musical performance.



OLIGODENDROCYTE (*green*) prepares to coat an axon (*purple*) with myelin.

Several labs have verified that action potentials, signals coursing the length of axons, stimulate myelination of these exposed areas of neural wiring. In 2014 Michelle Monje's lab at Stanford University showed that optogenetic stimulation (using lasers to make neurons fire) increased myelination in the mouse brain. That same year William Richardson's lab at University College London demonstrated that when the formation of new myelin is prevented, mice are slower to learn how to run on a wheel with some of its rungs removed. In studies where they used a confocal microscope to watch myelin form in live zebra fish, researchers in David Lyons's lab at the University of Edinburgh and in Bruce Appel's lab at the University of Colorado Denver observed that when the release of small sacs containing neurotransmitters from axons is inhibited, often the first few wraps of myelin slip off, and the oligodendrocyte aborts the entire process.

In 2018, working with our colleagues, including Daisuke Kato and others from various institutions in Japan, we showed how myelin promotes learning by ensuring that various spiking electrical signals traveling along axons arrive at the same time in the motor cortex, the brain region that controls movement. Using genetically modified mice with impaired myelination that had been trained to pull a lever to receive a reward, we found that learning this task increased myelination in the motor cortex.

By using electrodes to record neural impulses, we found that action potentials were less synchronized in the motor cortices of mice with faulty myelination. We then boosted the synchronization of spike arrivals in the motor cortex by using optogenetics to make neurons fire at the appropriate time. The mice with impaired myelination then performed the learned task proficiently. Eventually less invasive forms of brain stimulation may become effective therapy to treat neurological and psychological disorders caused by disrupted myelination.

Despite these recent advances, stimulation to increase axon myelination is not always enough to enable new learning, because we cannot synchronize the arrival of spikes at critical relay points in neural networks simply by making the impulses travel as rapidly as possible. There must also be a way to slow the speed of impulses from inputs that arrive at those points too soon.

The myelin that has already formed on axons has to be thickened or thinned in a controlled way to speed or slow signal transmission. Prior to our findings, there was no known explanation for how the myelin sheath could be thinned to slow signals, aside from disease damage. Our research revealed another type of glial cell involved in these "plastic" nervous system changes.

Surrounding the node of Ranvier is a glial cell called an astrocyte. Astrocytes have many functions, but most neuroscientists have largely ignored them because they do not communicate with other cells through electrical impulses. Surprisingly, research in the past decade has shown that astrocytes positioned close to the synapse between two neurons can regulate synaptic transmission during learning by releasing or taking up neurotransmitters there. But until fairly recently, myelin biologists tended to ignore the unique type of astrocyte that contacts an axon at a node of Ranvier.

What exactly do these so-called perinodal astrocytes do to thin the myelin sheath? Just as one would begin when remodeling a garment, these cells assist in cutting the "seams." The myelin sheath is attached to the axon by a spiral junction flanking the node of Ranvier. Under an electron microscope these junctions appear as spirals of stitches between the axon and the myelin, and the threads that form each stitch are composed of a complex of three cell adhesion molecules. Our analysis of the molecular composition of these stitch points showed that one of these molecules, neurofascin 155, has a site that can be cleaved by a specific enzyme, thrombin, to thin the myelin.

Thrombin is made by neurons, but it also can enter the brain from the vascular system. As the myelin lifts off the axon, the amount of bare axon at the node of Ranvier increases. The outer layer of myelin is attached to the axon adjacent to the perinodal astrocytes. When the myelin is detached from the axon, the outer layer withdraws into an oligodendrocyte, thinning the sheath. Both widening of the nodal gap and thinning of the myelin sheath slow the speed of impulse transmission.

We found that the enzyme's snipping of these threads that stitch myelin to the axon can be controlled by the perinodal astrocyte's release of an inhibitor of thrombin. We carried out experiments on genetically modified mice in which astrocytes released less of this thrombin inhibitor. When we looked at their neurons with an electron microscope, we could see that the myelin had thinned and that the nodal gap had increased. By using electronic

R. Douglas Fields/National Institutes of Health and NICHD

amplifiers to detect neural impulses and measure their speed of transmission, we found that after the myelin thickness decreased in this way, the speed of impulse transmission in the optic nerve slowed by about 20 percent and the animals' vision declined. We were able to reverse all these changes by injecting thrombin inhibitors, which are approved for treating vascular disorders.

Our experiments support a new hypothesis: the myelin sheath's changes in thickness represent a new form of nervous system plasticity governed by the addition and subtraction of myelin. Additional layers of myelin are not added to axons as one would wrap tape around a wire, because this would tie the legs of the oligodendrocytes in knots. Instead new insulation is affixed through the construction of a new inner layer that spirals around the axon like a snake below the overlying myelin. Meanwhile the outer layer of myelin can be detached by the perinodal astrocyte to thin the sheath. The thickness of the myelin sheath is not fixed; instead it reflects a dynamic balance between the addition of layers next to the axon and removal of the outer layer under control of the astrocyte.

BRAINY WAVES

THE OPTIMAL TIMING of action potentials at relay points is critical for strengthening synapses by adjusting their timing to allow them to fire together. But myelin plasticity can contribute to neural circuit function and learning in another way—by tuning the frequency of brain-wave oscillations. Not all neural activity in the brain arises from sensory inputs. Much of it takes place because of what goes on in the brain itself at both conscious and unconscious levels. This self-generated activity consists of oscillating waves of different frequencies that sweep through the brain, just as the vibration of a car engine at a certain speed will set different parts of the automobile rattling together at resonant frequencies.

These brain waves, or oscillations, are believed to be a key mechanism for coupling neurons across distant regions of the brain, which may be important for sorting and transmitting neural information. Oscillations, for example, tie together neural activity in the prefrontal cortex, which provides contextual meaning, and in the hippocampus (responsible for encoding spatial information). This oscillatory coupling enables a person to quickly recognize a familiar face at work, but it also makes it more difficult to identify the same co-worker in an unfamiliar place.

More important, the various stages of sleep, critical for storing long-term memories, can be identified by brain waves oscillating at different frequencies. Our experiences accumulated during the day are replayed during sleep and sorted for storage or deletion based on how they relate to other memories and emotions, which can mark them as potentially useful (or not) in the future. Appropriate brain-wave oscillations are believed to be pivotal in this process of memory consolidation. But the speed of impulse transmission is critical in synchronizing brain waves.

Just as two toddlers must precisely time their leg movements to drive the up-and-down motion of a teeter-totter, the transmission delays between two populations of oscillating neurons must be timed so that coupled neurons oscillate in synchrony across long distances in the brain. Myelin plasticity is important for brain waves because the proper conduction velocity is necessary to sustain oscillations that couple two regions of the brain at the same frequency.

This conclusion is based on mathematical modeling of the fundamental physics of wave propagation done by me, together with my NIH colleagues Sinisa Pajevic and Peter Basser. In 2020 a study by Patrick Steadman and his colleagues in Paul Frankland's lab at the University of Toronto provided convincing experimental support for the idea. Using genetically modified mice in which myelination could be temporarily halted, the researchers found that the ability to learn to fear an unsafe environment and to remember safe locations depends on the formation of new myelin. Moreover, they found that in this type of learning, brain-wave activity during sleep becomes coupled between the hippocampus and the prefrontal cortex. The prevention of new myelin formation also weakened connections and resulted in a type of impaired recall often found in people who have difficulty associating fear after a traumatic event with the appropriate context.

Learning and performing any complex task involves the coordinated operation of many different neurons in diverse brain regions and requires that signals proceed through large neural networks at an optimal speed. The myelin sheath is crucial for optimal transmission, but people begin to lose myelin in the cerebral cortex in their senior years. This gradual degradation is one of the reasons for cognitive slowing and the increasing difficulty of learning new things as we age.

Consider how transmission delays disrupt long-distance communication by telephone. Similarly, lags in the brain can cause cognitive difficulties and disorganized thinking in individuals with psychological disorders such as schizophrenia. Indeed, differences in brain-wave oscillations are seen in many neurological and psychiatric disorders. Alzheimer's disease, for instance, is associated with changes in white matter.

Drugs that control myelin production could provide new approaches to treating these problems. Indeed, Fei Wang and his colleagues at the Third Military Medical University in China, in collaboration with Jonah Chan of the University of California, San Francisco, reported in 2020 that the myelination-promoting drug clemastine given to mice with a gene deletion that impairs development of oligodendrocytes improved learning tested in a water maze. Because myelination is influenced by many forms of neural activity, a number of techniques—for example, cognitive training, neurofeedback and physical therapy—may be helpful in treating age-related cognitive decline and other disorders. A 2018 study of older adults by Jung-Hae Youn and his colleagues in South Korea indicated that 10 weeks of memory-training exercises increased recall. Brain imaging before and after training revealed increased integrity of white matter tracts connecting to the frontal lobe in the group of seniors who undertook the memory-training sessions.

These novel concepts have begun to change the way we think about how the brain works as a system. Myelin, long considered inert insulation on axons, is now seen as making a contribution to learning by controlling the speed at which signals travel along neural wiring. In venturing beyond the synapse, we are beginning to fill out the stick-figure skeleton of synaptic plasticity to create a fuller picture of what happens in our brain when we learn. SA

R. Douglas Fields is a senior investigator at the National Institutes of Health's Section on Nervous System Development and Plasticity. He is author of *Electric Brain: How the New Science of Brainwaves Reads Minds, Tells Us How We Learn, and Helps Us Change for the Better* (BenBella Books, 2020).

THE BRAIN'S SOCIAL ROAD MAPS

Neural circuits that track our whereabouts in space and time may also play vital roles in determining how we relate to other people

By Matthew Schafer and Daniela Schiller

Illustration by Richard Borge

WE ARE OFTEN TOLD THAT THERE ARE NO SHORTCUTS IN LIFE. But the brain—even the brain of a rat—is wired in a way that completely ignores this kind of advice. The organ, in fact, epitomizes a shortcut-finding machine.

The first indication that the brain has a knack for finding alternative routes was described in 1948 by Edward Tolman of the University of California, Berkeley. Tolman performed a curious experiment in which a hungry rat ran across an unpainted circular table into a dark, narrow corridor. The rat turned left, then right, and then took another right and scurried to the far end of a well-lit narrow strip, where, finally, a cup of food awaited. There were no choices to be made. The rat had to follow the one available winding path, and so it did, time and time again, for four days.

On the fifth day, as the rat once again ran straight across the table into the corridor, it hit a wall—the path was blocked. The animal went back to the table and started looking for alternatives. Overnight, the circular table had turned into a sunburst arena. Instead of one track, there were now 18 radial paths to explore, all branching off from the sides of the table. After venturing out a few inches on a few different paths, the rat finally chose to run all the way down path number six, the one leading directly to the food.

Taking the path straight to the food cup without prior experience may seem trivial, but from the perspective of behavioral psychologists at the time, the rat's navigational accomplishment was a remarkable feat. The main school of animal learning in that era believed that maze behavior in a rat is a matter of simple stimulus-response associations. When stimuli in the environment reliably produce a successful response, neural connections that represent this association get strengthened.

In this view, the brain operates like a telephone switchboard that maintains only reliable connections between incoming calls from our sense organs and outgoing messages to the muscles. But the behavioral switchboard was unable to explain the ability to correctly choose a shortcut right off the bat without having first experienced that specific path. Shortcuts and many other intriguing observations along these lines lent support to a rival school of thought promulgated by theorists who believe that in the course of learning, a map gets established in a rat's brain. Tolman—a proponent of that school—coined the term: the cognitive map.

According to Tolman, the brain does more than just learn the direct associations among stimuli. Indeed, such associations are often brittle, rendered outdated by changes in the environment. As psychologists have learned in the decades since Tolman's work, the brain also builds, stores and uses mental maps. These models of the world enable us to navigate our surroundings, despite complex, changing environments—affording the flexibility to use shortcuts or detours as needed. The hungry rat in Tolman's experiment must have remembered the location of the food, inferred the angle to it and chosen the route most likely to bring it to its goal. Quite simply, it must have built a model of the environment.

Such model building or mapmaking extends to more than physical space. Mental maps may exist at the core of many of our most "human" capacities, including memory, imagination, inferences, abstract reasoning and even the dynamics of social interactions. Researchers have begun to explore whether mental maps document how close or distant one individual is to another and where that individual resides in a group's social hierarchy. How does the brain, in fact, create the maps that allow us to make our way about the world?

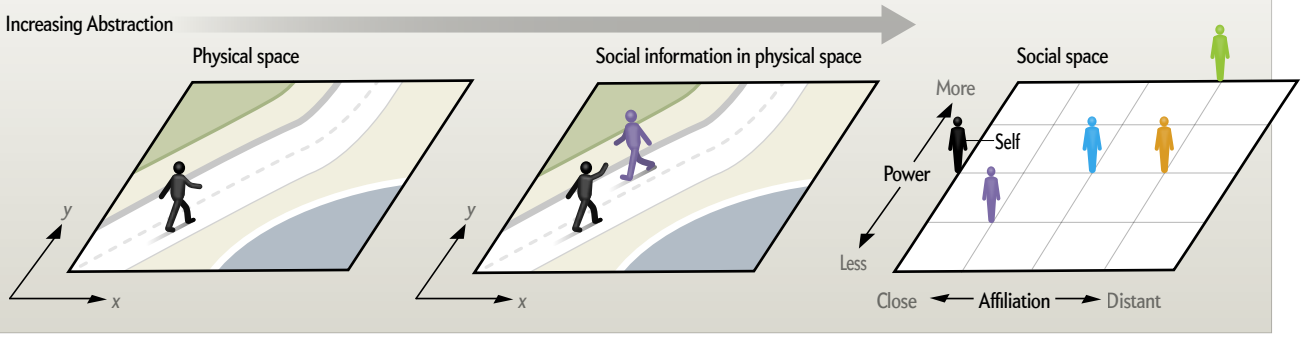
A SPATIAL MAP

THE FIRST HINTS of a neural basis for mental maps came in the 1970s. While studying a brain region called the hippocampus in rodents, John O'Keefe of University College London, along with his student Jonathan Dostrovsky, discovered a particular class of neurons that becomes active when mice occupy specific locations in their environment. Some of these neurons fired when the animal was in one location, and others switched on when it moved to the next spot on the path along which it traveled, as if the cells were specialized to track *where* the animal was in space. By linking sequences of these "place cells" together, researchers were able to reconstruct an animal's navigational trajectory. Work over the intervening decades confirmed the existence of place cells in other animals, including humans, and clarified many of their properties. Along the way, a host of cell types surfaced, each uniquely contributing to the brain's encoding of spatial representations.

In the nearby entorhinal cortex, a region connected to the hippocampus, a research team led by Edvard Moser and May-Britt Moser, former postdoctoral visiting fellows in O'Keefe's laboratory, discovered neurons highly similar to place cells. These cells also fired when an animal was in specific locations. But unlike place cells, each of these newly discovered cells spiked in multiple, regular locations. When mapped onto the animal's position, the activity patterns of these "grid cells" resembled highly regular, equilateral triangles. Like a spatial metric, these cells fired when an animal passed over the vertices of the triangles. The discovery of these cell types sparked excitement because of the emerging picture of how the brain controls navigation. Place cells and grid cells could provide a means to

Giving Way to the Abstract

Maps simplify the world by reducing an overwhelming amount of sensory and cognitive data into a format that can be used for navigating physical space, pointing to shortcuts and detours to reach a destination faster. The organization of such maps—built on the activity of cells dedicated to tracking both space and time—scales in the abstraction of what they represent: from the recognition of another individual along the way to even a complex space that denotes social power and closeness to others.



locate oneself in space and determine distance and direction. These navigational tools are crucial for building mental maps. (O'Keefe and the Mosers received the 2014 Nobel Prize in Medicine or Physiology for their work on place and grid cells.)

A wide variety of information is useful for creating such a map, and the hippocampus-entorhinal system encodes much of it. Discovering the location of a physical goal is one example: as an animal navigates toward an objective, some hippocampal neurons fire depending on the direction and distance to reach it. The cells increase their firing rate as the animal approaches the goal.

Other cells also enter the picture. A dedicated population of "reward" cells encodes reward locations across different environments, providing a signal to guide an animal's navigation (think of an "X" marking the spot of treasure on a pirate's map). Other cells track speed and direction and in doing so act like internal speedometers and compasses that compute an animal's progress as it travels through the environment. Specific cells that signal the locations of landmarks in the surroundings serve as references to correct errors in the animal's trajectory. A map must also have edges: cells that fire more as the animal approaches the map's perimeter.

For humans, the importance of such an abundance of cell types seems obvious: the brain is responsible for knowing the location of home and work, walls and dead ends, a favorite shop or the corner store. It is still a mystery as to how all of this information is drawn together into a coherent map, but these cells appear to provide the parts list for the elements of neural mapmaking.

This hippocampal-entorhinal system is more than a mapmaker, though, and the maps are more than a way to locate oneself in space. These maps also are used for active planning. When a rat comes to a junction in a familiar maze, it will pause while place cell firing sequences that relate to the different options are activated, as if the animal is contemplating the choices.

Humans engage similar processes. Research in participants navigating virtual environments while their brains were scanned with functional magnetic resonance imaging shows that the hippocampus becomes active in ways consistent with spatial planning, such as considering and planning routes.

Shaping plans also occurs during sleep. Sequences of place cell activity can be reactivated during sleep to replay the past or simulate the future. Without the ability to simulate new behaviors, we would have to explore a multitude of real-world options before deciding on what action to take. We would be constant empiricists, able to act only on direct observations. Instead offline simulations give us the ability to envision possibilities without directly experiencing them.

MENTAL TIME TRAVEL

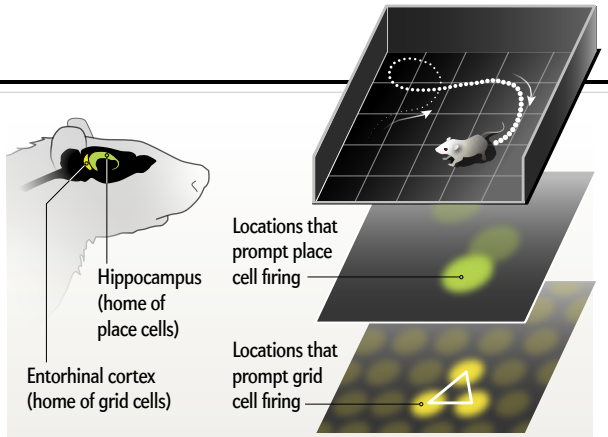
TIME AND SPACE are inextricably linked. It is difficult to talk about time without borrowing a spatial metaphor: time "passes" as we "move" through it. We look "forward" to the future and "back" on our past. The same hippocampal-entorhinal system tracks movement through time. Work done largely in the lab of the late Howard Eichenbaum of Boston University revealed neurons in the hippocampal-entorhinal system that encode the time course of an animal's experience. Time cells fire at successive moments but do not track time in a simple clocklike fashion. Instead they mark temporal context—stretching or shrinking their firing durations if the length of a task changes, for example. Some time cells encode space as well. In the brain, in fact, physical and temporal space may be bound together.

The discovery of the crucial importance of these brain areas in space and time was not totally surprising. Psychologists had long suspected it to be the case. In 1953 Henry Molaison underwent bilateral hippocampal resection surgery to reduce extreme, life-disrupting epileptic seizures. The surgery was successful at quelling the seizures. But Molaison—known for decades only as H.M.—became one of the most renowned cases in the history of the brain sciences.

Molaison could remember most experiences from before his surgery—people he knew and recollections from culture and politics. But his ability to form such explicit memories postsurgery was practically nonexistent. Even so, certain types of learning and memory remained untouched: he could still learn some new skills with enough practice. But his recollections of new people, facts and events were immediately lost.

Source: "Navigating Social Space," by Matthew Schafer and Daniela Schiller, in *Neuron*, Vol. 100, October 24, 2018

Illustrations by Jen Christiansen

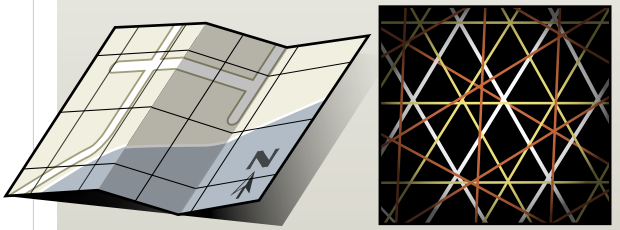


Cognitive Cartography Is Physical *and* Social

The brain forms the idea of friend or foe by stitching together diverse social characteristics from memories that track one's whereabouts. The recollections, research suggests, can then be used to place an individual within a social hierarchy that elucidates, say, where one stands in relation to others.

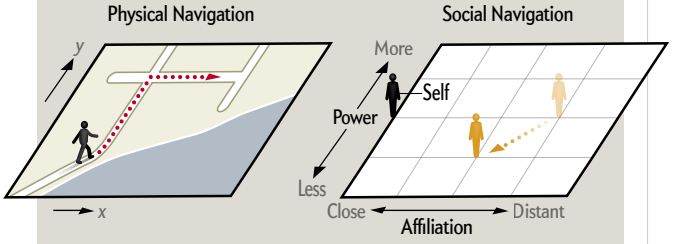
PLACE AND GRID CELLS

Place cells pinpoint the animal's whereabouts, each cell firing when a particular spot on a mental map is reached. A grid cell activates when an animal passes over the vertices of triangles superimposed on a mental map. The triangles' pattern of activation helps the animal compute the direction and distance traveled along a route.



MAKING THE LEAP TO SOCIAL MAPS

Go right at the corner and continue to your destination. Building a map of physical surroundings is the work of place and grid cells. But the brain may also use these cells for constructing maps for social milieus: locating an acquaintance who grows closer but loses power in a relationship.



From observing Molaison, neuroscientists discerned that the hippocampus was essential in forming the episodic memories that record facts and events. Research on the role of the hippocampus in episodic memory exploded, largely in parallel to studies on its maplike functions.

The discoveries about the roles of the hippocampus and entorhinal cortex in spatial navigation and episodic memory were significant for at least a couple of reasons. The work in spatial navigation in rodents marked the first time that a higher-order cognitive function—something beyond basic sensory processes—mapped onto clear neural correlates. H.M. showed us that there were multiple types of memory supported by at least partially different neural systems, with the hippocampus playing a central role in the formation and storage of new episodic memories. These discoveries hinted that mechanisms of spatial and temporal navigation might underlie episodic memory. This synthesis is perhaps best explained by the theoretical construct proposed decades earlier by Tolman; both episodic memory and spatial navigation might reflect the brain's formation and use of cognitive maps.

Maps are not accurate portraits of the world in all of its complexity. Rather they are representations of relations—distances and directions between locations and what exists where. Maps reduce a dizzying amount of real-world information into a simple, easily readable format that is useful for effective, flexible navigation. The cell types mentioned earlier (place cells, grid cells and border cells, among others) may piece together such related elements into a mental map, which other brain regions can then read out to guide "navigation," amounting to adaptive decision-making. Mapping allows relations to be inferred, even when they have not been experienced. It also allows for mental shortcuts that go beyond the purview of the spatial and temporal domains. In fact, reasoning using abstract concepts may depend on some of these same neural foundations.

In one example of this new line of work, researchers Alexandra Constantinescu, Jill O'Reilly and Timothy Behrens, all then at the University of Oxford, asked participants to learn associations of different symbols with images of "stick" birds with various neck and leg lengths. A bird with a long neck but short legs, for example, might be linked with the image of a bell, whereas a bird with a short neck and long legs might be connected to a teddy bear. These linkages created a two-dimensional association space. Despite neuroimaging being too crude to detect actual grid cells in the human brain, imaging conducted during the learned-association testing nonetheless revealed a gridlike pattern of activation within the entorhinal cortex.

This finding builds on earlier work by Christian Doeller of the Max Planck Institute for Human Cognitive and Brain Science in Leipzig, Germany, and Neil Burgess of University College London that first showed an entorhinal gridlike representation in humans navigating a virtual maze. For both physical and abstract relations, the gridlike organization is highly efficient. It makes the linkages of places or concepts more predictable, enhancing how quickly inferences can be made about these relations. As in physical space, this organization of information allows for inferring shortcuts—relations between ideas or perhaps analogies, stereotypes and even some aspects of creativity itself could depend on such inferences.

Sources: "Scientific Background: The Brain's Navigational Place and Grid Cell System," by Ole Kiehn and Hans Forssberg, with illustrations by Mattias Karlen. Nobelprize.org; "Navigating Social Space," by Matthew Schafer and Daniela Schiller, in *Neuron*, Vol. 100; October 24, 2018

PEOPLE MAPS

THE PROGRESSION from the physical to the abstract carries over into the way the brain represents social relationships. Various bits of knowledge about another person are distilled into the concept of that individual. When we see a photograph of someone or hear or see that person's name, the same hippocampal cells will fire, regardless of the sensory details of the stimulus (for example, the famous "Jennifer Aniston neuron" described by Itzhak Fried of the University of California, Los Angeles, and his colleagues). These hippocampal cells are responsible for representing concepts of specific individuals.

Other hippocampal cells track the physical locations of others and are called social place cells. In an experiment by David Omer of the Hebrew University of Jerusalem, Nachum Ulanovsky of the Weizmann Institute of Science in Rehovot, Israel, and their colleagues, bats observed other bats navigating a simple maze to reach a reward. The task of an observer bat was to simply watch and learn from a navigating bat, enabling it to subsequently navigate the same route to get the same reward. When the observer bat watched, hippocampal cells fired corresponding to the location of the other bat.

Neural circuitry within specific subregions of the hippocampus (in particular, areas called CA1 and CA2) contributes to such social memories. Artificial stimulation or inactivation of these hippocampal areas enhances or diminishes an animal's ability to recognize other animals. In humans, hippocampal injury often spares memory for specific, individual faces, but the relation between this cardinal identifier of another person and that individual's behavior may be lost. That observation suggests that the hippocampus does not simply record a face or some other personal detail but rather ties together diverse social characteristics.

Hippocampal activity also tracks social hierarchies: the demands of a boss and a co-worker, for instance, may be valued differently and confer different social standings. Common metaphors illustrate the spatial dimensions of a hierarchy: a person may try to gain status to "climb the social ladder" or "look down" at someone below them. Other factors are also critical. Biological relatedness, common group goals, the remembered history of favors and slights—all determine social proximity or distance. Human relationships can be conceived of as geometric coordinates in social space that are defined by the dimensions of hierarchy and affiliation.

Work in our lab has explored these ideas in recent years. Our results suggest that, as with other spaces, the hippocampus organizes social information into a maplike format. To test this hypothesis, we put individuals in a choose-your-own-adventure game in which they interacted with cartoon characters and made decisions while their brains were scanned.

In the game, players had just moved to a new town and needed to interact with the fictional characters to secure a job and a place to stay. Participants made decisions on how to deal with a given character. Players could request that others perform favors to demonstrate their power, or they could submit to demands made on them. In a subsequent interaction, they could decide whether or not to make a gesture of attachment—giving a hug or remaining at a distance.

Using these decisions, we plotted each character at certain coordinates on a map representing their movement along the dimensions of power and affiliation. In each interaction, we drew a line or vector from the participant to the character. In this way, we charted the evolving relations as trajectories through social space and computed information about the angles and lengths of the social vectors.

We searched for neural signals that tracked this information by correlating a participant's brain activity with the angle and length of the vectors for each decision. Activity in the hippocampus tracked the angle of the characters to the participant. The degree to which hippocampal activity captured these social coordinates also reflected the participants' self-reported social skills. These findings suggest that the hippocampus monitors social dynamics as it does physical locations by encoding relations between points in multidimensional space. Indeed, it may be that along any arbitrary dimension in which we can order information, whether physical or abstract, the hippocampus-entorhinal system plays a part.

Many questions about the brain's social maps still remain unanswered. How does this system interact with other brain regions? For example, in our role-playing study, we found that the posterior cingulate cortex, a region also involved in representing spatial information, tracked the length of social vectors—functioning in effect as a measuring stick of "social distance." Further, a gridlike signal was found in brain regions that are interconnected with and tend to co-activate with the hippocampal-entorhinal system, suggesting they form a network of brain regions with common functional properties.

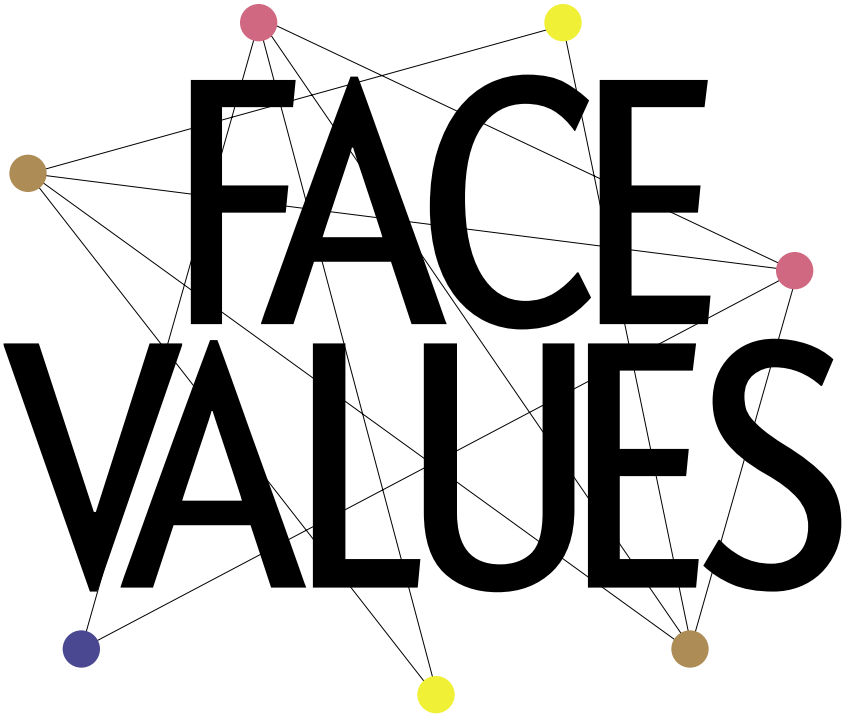
As research accumulates, questions of clinical importance arise as well. Can flawed mapping processes explain psychiatric dysfunction? Another possibility is that insights garnered from this brain architecture could inform artificial-intelligence development. Well-organized internal models of the world might be key to building more intelligent machines.

That the same mapping system may underlie navigation through space and time, reasoning, memory, imagination and even social dynamics suggests that our ability to construct models of the world might be what makes us such adaptive learners. The world is full of both physical and abstract relations. Road maps of city streets and mental maps of interrelated concepts help us make sense of the world by extracting, organizing and storing related information. A new coffee shop on a familiar street can be easily placed within an existing spatial map. Fresh concepts can be related to older ideas. And a new acquaintance can reshape our social space.

Maps let us simulate possibilities and make predictions, all within the safety of our own heads. The mental shortcuts we can so readily conjure up might have their basis in the same system that allows us to figure out a detour around a traffic jam. We have just begun to discover the varied properties and capacities of this system. Mental maps do more than help us find shortcuts through physical space—they enable us to navigate life itself. SA

Matthew Schafer is pursuing a doctorate in neuroscience at the Icahn School of Medicine at Mount Sinai, focusing on the neural mechanisms of social cognition in the human brain.

Daniela Schiller is an associate professor of both neuroscience and psychiatry at the Icahn School of Medicine at Mount Sinai. She researches the neural mechanisms underlying emotional control needed to adapt to constantly changing environments.



Brain regions that process faces reveal deep insights into the neural mechanisms of vision

By Doris Y. Tsao

Illustration by Brian Stauffer

WHEN I WAS IN HIGH SCHOOL, I LEARNED ONE DAY ABOUT THE DENSITY of curves in an introductory course on calculus. A simple pair of differential equations that model the interactions of predators and prey can give rise to an infinite number of closed curves—picture concentric circles, one nested within another, like a bull's-eye. What is more, the density of these curves varies depending on their location.

This last fact seemed so strange to me. I could easily imagine a finite set of curves coming close together or pulling apart. But how could an infinity of curves be denser in one region and less dense in another? I soon learned that there are different types of infinity with paradoxical qualities, such as Hilbert's Hotel (where the rooms are always fully booked but new guests can always be accommodated) and the Banach-Tarski apple (which can be split into five pieces and rearranged to make two apples with the same volume as the original). I spent hours poring over these mathematical proofs. Ultimately they struck me as symbolic magic of no real consequence, but the seed of interest had taken root.

Later, as an undergraduate at the California Institute of Technology, I learned about the experiments of David Hubel and Torsten Wiesel and their landmark discovery of how a region in the brain called the primary visual cortex extracts edges from the images relayed from the eyes. I realized that what had mystified me back in high school was the act of trying to *imagine* different densities of infinity. Unlike the mathematical tricks I had studied in high school, the edges that Hubel and Wiesel described are processed by neurons, so they actually exist in the brain. I came to recognize that visual neuroscience was a way to understand how this neural activity gives rise to the conscious perception of a curve.

The sense of excitement this realization triggered is hard to describe. I believe at each stage in life one has a duty. And the duty of a college student is to dream, to find the thing that captures one's heart and seems worth devoting a whole life to. Indeed, this is the single most important step in science—to find the right problem. I was captivated by the challenge of understanding vision and embarked on a quest to learn how patterns of electrical activity in the brain are able to encode perceptions of visual objects—not just lines and curves but even objects as hard to define as faces. Accomplishing this objective required pinpointing the specific brain regions dedicated to facial recognition and deciphering their underlying neural code—the means by which a pattern of electrical impulses allows us to identify people around us.

FACE PATCHES DO ACT AS AN ASSEMBLY LINE TO SOLVE ONE OF THE BIG CHALLENGES OF VISION: HOW TO RECOGNIZE THINGS AROUND US DESPITE CHANGES IN THE WAY THEY LOOK.

The journey of discovery began in graduate school at Harvard University, where I studied stereopsis, the mechanism by which depth perception arises from differences between the images in the two eyes. One day I came across a paper by neuroscientist Nancy Kanwisher, now at the Massachusetts Institute of Technology, and her colleagues, reporting the discovery of an area in the human brain that responded much more strongly to pictures of faces than to images of any other object when a person was inside a functional magnetic resonance imaging (fMRI) brain scanner. The paper seemed bizarre. I was used to the brain being made of parts with names like "basal ganglia" and "orbitofrontal cortex" that had some vague purpose one could only begin to fathom. The concept of an area specifically devoted to processing faces seemed all too comprehensible and therefore impossible. Anyone could make a reasonable conjecture about the function of a face area—it should probably represent all the different faces that we know and something about their expression and gender.

As a graduate student, I had used fMRI on monkeys to identify areas activated by the perception of three-dimensionality in images. I decided to show pictures of faces and other objects to a monkey. When I compared activation in the monkey's brain in response to faces with activation for other objects, I found several areas that lit up selectively for faces in the temporal lobe (the area underneath the temple)—specifically in a region called the inferotemporal (IT) cortex. Charles Gross, a pioneer in the field of object vision, had discovered face-selective neurons in the IT cortex of macaques in the early 1970s. But he had reported that these cells were randomly scattered throughout the IT cortex. Our fMRI results provided the first indication that face cells might be concentrated in defined regions.

FACE PATCHES

AFTER PUBLISHING MY WORK, I was invited to give a talk describing the fMRI study as a candidate for a faculty position at Caltech, but I was not offered the job. Many people were skeptical of the value of fMRI, which measures local blood flow, the brain's plumbing. They argued that showing increased blood flow to a brain area when a subject is looking at faces falls far short of clarifying what neurons in the area are actually encoding because the relation between blood flow and electrical activity is unclear. Perhaps by chance these face patches simply contained a slightly larger number of neurons responsive to faces, like icebergs randomly clustered at sea.

Because I had done the imaging experiment in a monkey, I could directly address this concern by inserting an electrode into an fMRI-identified face area and asking, What images drive single neurons in this region most strongly? I performed this experiment together with Winrich Freiwald, then a postdoctoral fellow in Margaret Livingstone's laboratory at Harvard, where I was a graduate student. We presented faces and other objects to a monkey while amplifying the electrical activity of individual neurons recorded by the electrode. To monitor responses in real time, we converted the neurons' electrical signals to an audio signal that we could hear with a loudspeaker in the lab.

This experiment revealed an astonishing result: almost every single cell in the area identified through fMRI was dedicated to processing faces. I can recall the excitement of our first recording, hearing the "pop" of cell after cell responding strongly to faces and very little to other objects. We sensed we were on to something important, a piece of cortex that could reveal the brain's high-level code for visual objects. Marge remarked on the face patches: "You've found a golden egg."

I also remember feeling surprised during that first experiment. I had expected the face area would contain cells that responded selectively to specific individuals, analogous to orientation-selective cells in the primary visual cortex that each respond to a specific edge orientation. In fact, a number of well-publicized studies had suggested that single neurons can be remarkably selec-

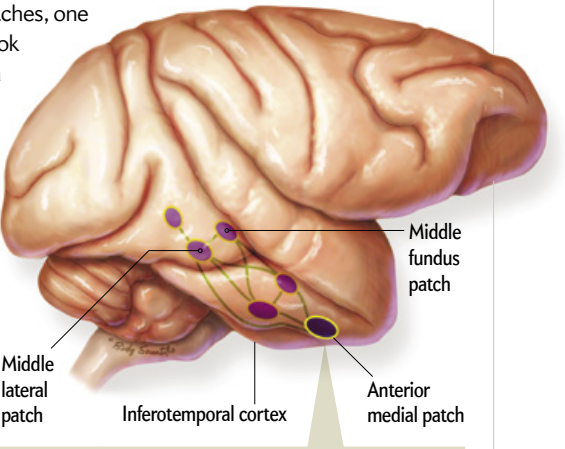
tive for the faces of familiar people—responding, say, only to Jennifer Aniston. Contrary to my expectation, each cell seemed to fire vigorously for almost any face.

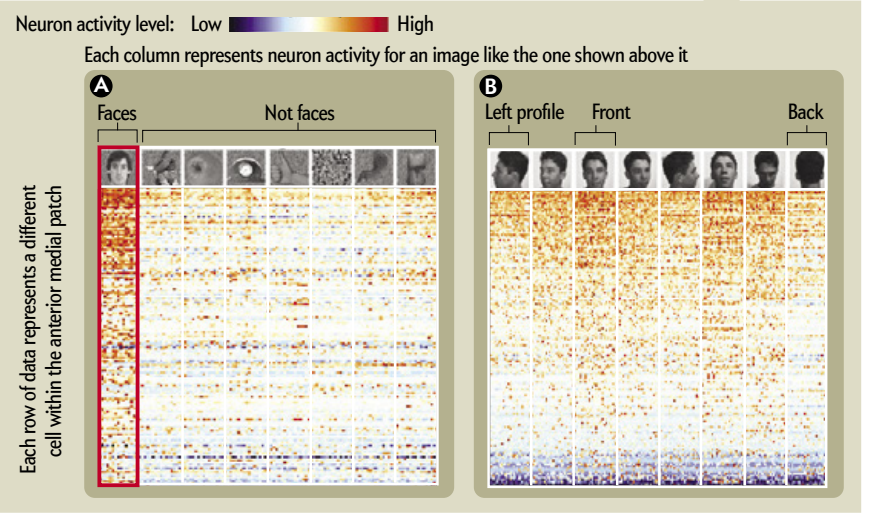
I plugged madly away at Photoshop during these early experiments and found that the cells responded not just to faces of humans and monkeys but even to highly simplified cartoon faces.

Observing this phenomenon, I decided to create cartoon faces with 19 different features that seemed pertinent to defining the identity of a face, including inter-eye distance, face aspect ratio and mouth height, among other characteristics. We then went on to alter these values—the inter-eye distance, for instance, varied from almost cyclopean to just inside the face boundary. Individual cells responded to most faces but interestingly did not always exhibit the exact same rate of firing with all faces. Instead there was a systematic variation in their response: when we plotted the firing of cells for the different cartoon features, we found a pattern in which there was a minimal response to one feature extreme—the smallest inter-eye distance, for instance—and a maximal response to the opposite extreme—the largest eye separation—with intermediate responses to feature values in the middle. The response as a function of the value for each feature looked like a ramp, a line slanted up or down.

Where Are the Face Detectors?

A set of six nodes in the inferotemporal cortex of both brain hemispheres specializes in identifying faces. These "face patches" function as an assembly line: in the middle lateral and middle fundus patches, one neuron might become active when faces look straight ahead; another might turn on for a face looking to the right. At the end of the assembly line, in the anterior medial patch, varying views are stitched together. Neurons in this patch are active in response to the face of a specific individual, no matter if the view is from the front or side. Responses from a face patch of one monkey are generated for faces but not objects (*red areas in* A) and for the same individual, such as the dark-haired man, from varying angles (*red areas in* B).




Neuron activity level: Low High
Each column represents neuron activity for an image like the one shown above it
A
Faces
Not faces
B
Left profile
Front
Back
Each row of data represents a different cell within the anterior medial patch

Once again, I was invited to give a job talk at Caltech. Returning, I had more to offer than just fMRI images. With the addition of the new results from single-cell recordings, it was clear to everyone that these face patches were real and likely played an important role in facial recognition. Furthermore, understanding their underlying neural processes seemed like an effective way to gain traction on the general problem of how the brain represents visual objects. This time I was offered the job.

CONTRAST IS KEY

AT CALTECH, my colleagues and I dug deeper into the question of how these cells detect faces. We took inspiration from a paper by Pawan Sinha, a vision and computational neuroscientist at M.I.T., that suggested faces could be discerned on the basis of specific contrast relations between different regions of the face—whether the forehead region is brighter than the mouth region, for example. Sinha suggested a clever way to determine *which* contrast relations can be used to recognize a face: they should be the ones that are immune to changes in lighting. For example, "left eye darker than nose" is a useful feature for detecting a face because it does not matter if a face is photographed with lighting from above, left, right or below: the left eye is *always* darker than the nose (check for yourself).

From a theoretical standpoint, this idea provides a simple, elegant computational mechanism for facial recognition, and we wondered whether face cells might be using it. When we measured the response of cells to faces in which different regions varied in brightness, we found that cells often had a significant preference for a particular contrast feature in an image.

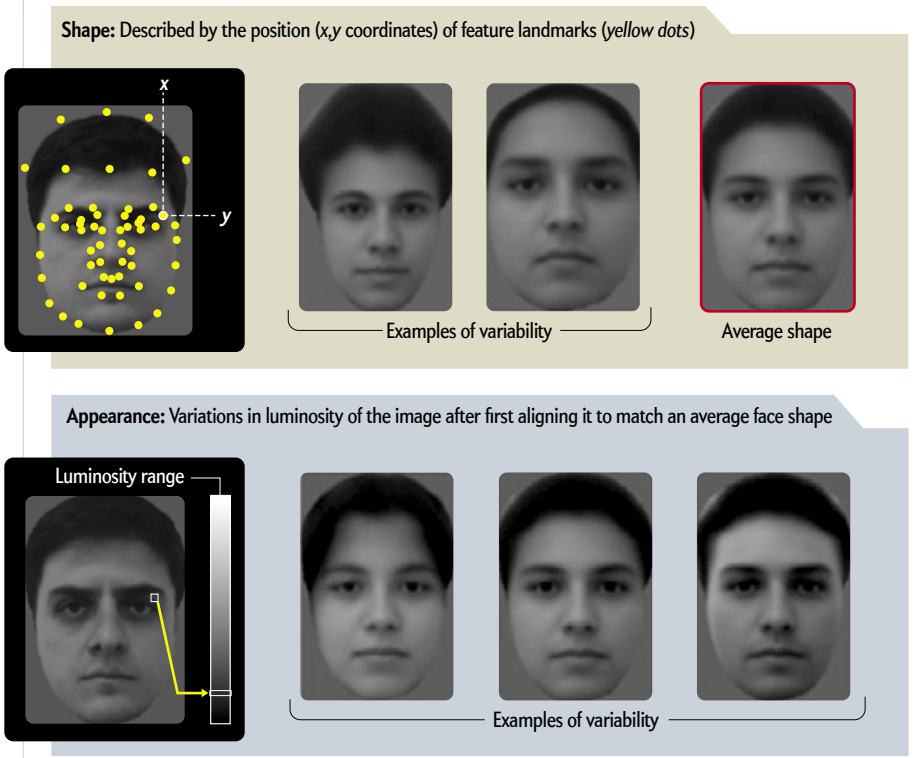
To our astonishment, almost all the cells were wholly consistent in their contrast preferences—just a single cell was found that preferred the opposite polarity. Moreover, the preferred features were precisely those identified by Sinha as being invulnerable to lighting

From "Functional Compartmentalization and Viewpoint Generalization within the Macaque Face-Processing System," by Winrich A. Freiwald and Doris Y. Tsao, in *Science*, Vol. 330, November 5, 2010 (*data grids* and *photo insets*)

Illustration by Body Scientific (brain)

Shape + Appearance = Face

Identifying the face patches was only a first step. It then became necessary to explore what happens in the neurons within each patch, setting off a search for the brain's coding scheme for faces. To derive quantitative measures for faces, the Tsao laboratory came up with 25 features for shape and 25 for appearance that could be used by each neuron in a face patch—a 50-dimensional face space. The shape features can be thought of as those defining the skeleton—how wide the head is or the distance between the eyes. The appearance features specify the face's surface texture (complexion, eye or hair color, and so on).


Shape: Described by the position (*x,y* coordinates) of feature landmarks (*yellow dots*)
x
y
Examples of variability
Average shape
Appearance: Variations in luminosity of the image after first aligning it to match an average face shape
Luminosity range
Examples of variability

changes. The experiment thus confirmed that face cells use contrast relations to detect faces.

More broadly, the result confirmed that these cells truly were face cells. At talks, skeptics would ask, How do you know? You can't test every possible stimulus. How can you be sure it's a face cell and not a pomegranate cell or a lawn mower cell? This result nailed it for me. The precise match between the way cells reacted to changes in contrast between different parts of the face and Sinha's computational prediction was uncanny.

Our initial experiments had revealed two nearby cortical patches that lit up for faces. But after further scanning (with the help of a contrast agent that increased severalfold the robustness of the signal), it became clear that there are in fact six face patches in each of the brain's two hemispheres (making a dozen golden eggs total). They are distributed along the entire length of the temporal lobe. These six patches, moreover, are not randomly scattered throughout the IT cortex. They are located in similar locations across hemispheres in each animal. Work by our group and others has found that a similar pattern of multiple face patches spanning the IT cortex exists in humans and other primates such as marmosets.

This observation of a stereotyped pattern suggested that the patches might constitute a kind of assembly line for processing faces. If so, one would expect the six patches to be connected to one another and each patch to serve a distinct function.

To explore the neural connections among patches, we electrically stimulated different patches with tiny amounts of current—a technique called microstimulation—while the monkey was inside an fMRI scanner. The goal was to find out what other parts of the brain light up when a particular face patch is stimulated. We discovered that whenever we stimulated one face patch, the other patches would light up, but the surrounding cortex would not, indicating that, indeed, the face patches are strongly interconnected. Furthermore, we found that each patch performs a different function. We presented pictures of 25 people, each at eight different head orientations, to monkeys and recorded responses from cells in three regions: the middle lateral and middle fundus patches (ML/MF), the anterior lateral patch (AL) and the anterior medial patch (AM).

We found striking differences among these three regions. In ML/MF, cells responded selectively to specific views. For example, one cell might prefer faces looking straight ahead, whereas another might opt for faces looking to the left. In AL, cells were less view-specific. One class of cells responded to faces looking up, down and straight ahead; another responded to faces looking to the left or right. In AM, cells responded to specific individuals regardless of whether the view of the face was frontal or in profile. Thus, at the end of the network in AM, view-specific representations were successfully stitched into a view-invariant one.

Apparently face patches do act as an assembly line to solve one of the big challenges of vision: how to recognize things around us despite changes in the way they look. A car can have any make and color, appear at any viewing angle and distance, and be partially obscured by closer objects such as trees or other cars. Recognizing an object despite these visual transformations is called the invariance problem, and it became clear to us that a major function of the face-patch network

From "The Code for Facial Identity in the Primate Brain," by Le Chang and Doris Y. Tsao, in *Cell*, Vol. 169, No. 6, June 1, 2017 (*face images*)

Graphics by Jen Christiansen

is to overcome this impediment.

Given the great sensitivity of cells in face patches to changes in facial identity, one might expect that altering these cells' responses should modify an animal's perception of facial identity. Neuroscientists Josef Parvizi and Kalanit Grill-Spector of Stanford University had electrically stimulated a face-patch area in human subjects who had electrodes implanted in their brains for the purpose of identifying the source of epileptic seizures and found that stimulation distorted the subjects' perception of a face.

We wondered whether we would find the same effect in monkeys when we stimulated their face patches. Would doing so alter the perception only of faces, or would it affect that of other objects as well? The boundary between a face and a nonface object is fluid—one can see a face in a cloud or an electrical outlet if prompted. We wanted to use electrical microstimulation as a tool to delineate precisely what constitutes a face for a face patch. We trained monkeys to report whether two sequentially presented faces were the same or different. Consistent with the earlier results in humans, we found that microstimulation of face patches strongly distorted perception so that the animal would always signal two identical faces as being different.

Interestingly, microstimulation had no effect on the perception of many nonface objects, but it did significantly affect responses to a few objects whose shape is consistent with a face—apples, for one. But why does this stimulation influence the perception of an apple?

One possibility is that the face patches are typically used to represent not just faces but also other round objects like apples. Another hypothesis is that face patches are not normally used to represent these objects, but stimulation induces an apple to appear face-like. It remains unclear whether face patches are useful for detecting any nonface objects.

CRACKING THE CODE

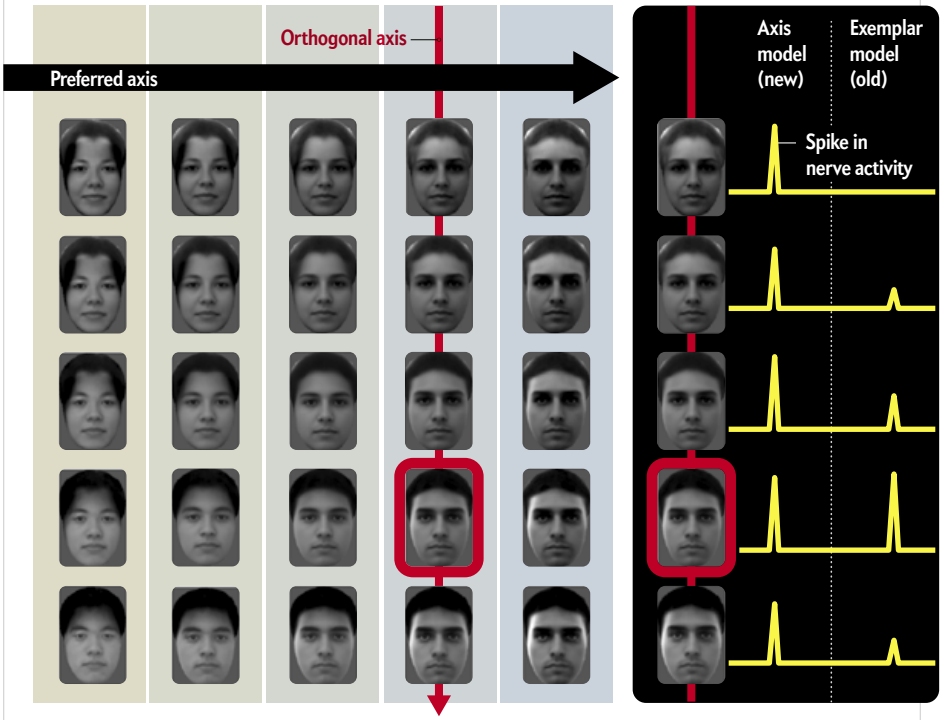
UNCOVERING the organization of the face-patch system and properties of the cells within was a major accomplishment. But my dream when we first began recording from face patches was to achieve something more. I had intuited that these cells would allow us to crack the neural code for facial identity. That means understanding how individual neurons process faces at a level of detail that would let us predict a cell's response to any given face or decode the identity of an arbitrary face based only on neural activity.

The central challenge was to figure out a way to describe faces quantitatively with high precision. Le Chang, then a postdoc in my lab, had the brilliant insight to adopt a technique from the field of computer vision called the active appearance model. In this approach, a face has two sets of descriptors, one for shape and another for appearance. Think of the shape features as those defined by the skeleton—how wide the head is or the distance between the eyes. The appearance features define the surface texture of the face (complexion, eye or hair color, and so on).

To generate these shape and appearance descriptors for faces, we started with a large database of face images. For each face, we placed a set of markers on key features. The spatial locations of these markers described

The Face Code, at Last

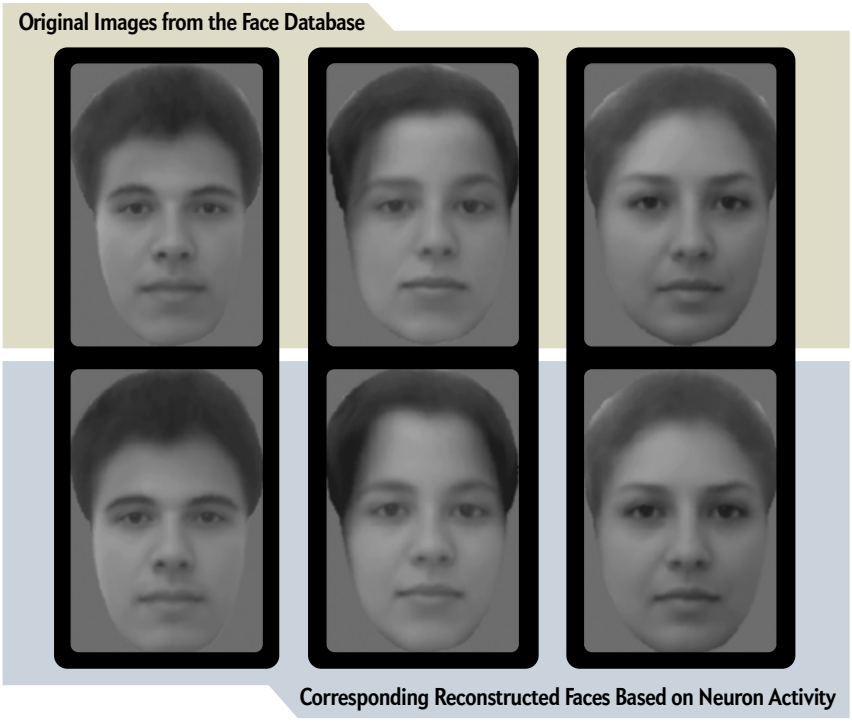
Having 50 coordinates that describe shape and appearance allows for a description of neurons' firing in response to a particular face—a description that functions as a code that can be visualized geometrically. In this code, each face cell receives inputs for a face in the form of the 50 coordinates, or dimensions. The neuron then fires with a particular intensity in response to a certain face (*red outlines*), along what is called the preferred axis. The intensity increases steadily (monotonically) along the preferred axis. Furthermore, the response is the same for every face on an axis at right angles to the preferred axis, even though those faces may look very different. This axis model of facial coding differs from a previous exemplar model that suggests that each neuron fires with maximum intensity to a single most preferred face.



From "The Code for Facial Identity in the Primate Brain," by Le Chang and Doris Y. Tsao, in *Cell*, Vol. 169, No. 6, June 1, 2017 (*face grid*)

Pictures Worth 205 Neurons

For a given face, we can predict how a cell will respond by taking a weighted sum of all 50 face coordinates. To predict what face the monkey saw from neuronal activity, this entire process can be reversed: by knowing the response of 205 face cells, it is possible to predict the 50 coordinates defining the exact facial features—and make a highly accurate reconstruction of a given face.



the shape of the face. From these varied shapes, we calculated an average face. We then morphed each face image in the database so its key features exactly matched those of the average face. The resulting images constituted the appearance of the faces independent of shape.

We then performed principal components analysis independently on the shape and appearance descriptors across the entire set of faces. This is a mathematical technique that finds the dimensions that vary the most in a complex data set.

By taking the top 25 principal components for shape and the top 25 for appearance, we created a 50-dimensional face space. This space is similar to our familiar 3-D space, but each point represents a face rather than a spatial location, and it comprises much more than just three dimensions. For 3-D space, any point can be described by three coordinates (x,y,z). For a 50-D face space, any point can be described by 50 coordinates.

In our experiment, we randomly drew 2,000 faces and presented them to a monkey while recording cells from two face patches. We found that almost every cell showed graded responses—resembling a ramp slanting up or down—to a subset of the 50 features, consistent with my earlier experiments with cartoon faces. But we had a new insight about why this is important. If a face cell has ramp-shaped tuning to different features, its response can be roughly approximated by a simple weighted sum of the facial features, with weights determined by the slopes of the ramp-shaped tuning functions. In other words:

response of face cells = weight matrix × 50 face features

We can then simply invert this equation to convert it to a form that lets us *predict* the face being shown from face cell responses:

50 face features = (1/weight matrix) × response of face cells

At first, this equation seemed impossibly simple to us. To test it, we used responses to all but one of the 2,000 faces to learn the weight matrix and then tried to predict the 50 face features of the excluded face. Astonishingly, the prediction turned out to be almost indistinguishable from the actual face.

A WIN-WIN BET

AT A MEETING in Ascona, Switzerland, I presented our findings on how we could reconstruct faces using neural activity. After my talk, Rodrigo Quian Quiroga, who discovered the famous Jennifer Aniston cell in the human medial temporal lobe in 2005 and is now at the University of Leicester in England, asked me how my cells related to his concept that single neurons react to the faces of specific people. The Jennifer Aniston cell, also known as a grandmother cell, is a putative type of neuron that switches on in response to the face of a recognizable person—a celebrity or a close relative.

I told Rodrigo I thought our cells could be the building blocks for his cells, without thinking very deeply about how this would work. That night, sleepless from jet lag, I recognized a major difference between our face cells and his. I had described in my talk how our face cells computed their response to weighted sums of different face features. In the middle of the night, I realized this computation is the same as a mathematical operation known as the dot product, whose geometric representation is the projection of a vector onto an axis (like the sun projecting the shadow of a flagpole onto the ground).

Remembering my high school linear algebra, I realized this implied that we should be able to construct a large "null space" of faces for each cell—a series of faces of varying identity that lie on an axis perpendicular to the axis of projection. Moreover, all these faces

Doris Y. Tsao (*face images*)

would cause the cell to fire in exactly the same way.

And this, in turn, would suggest cells in face patches are fundamentally different from grandmother cells. It would demolish the vague intuition everyone shared about face cells—that they should be tuned to specific faces.

I was the first person in the meeting's breakfast hall at 5 A.M. the next morning and hoped to find Rodrigo so I could tell him about this counterintuitive prediction. Amazingly, when he finally showed up, he told me he had the exact same idea. So we made a bet, and Rodrigo allowed the terms to be framed in a way that would be win-win for me. If each cell really turned out to have the same response to different faces, then I would send Rodrigo an expensive bottle of wine. If, on the other hand, the prediction did not pan out, he would send me solace wine.

In search of an answer back in our lab at Caltech, Le Chang first mapped the preferred axis for a given cell using responses to the 2,000 faces. Then he generated, while still recording from the same cell, a range of faces that could all be placed on an axis perpendicular to the cell's preferred axis. Remarkably, all these faces elicited exactly the same response in the cell. The next week Rodrigo received an exquisite bottle of Cabernet.

The finding proved that face cells are not encoding the identities of specific individuals in the IT cortex. Instead they are performing an axis projection, a much more abstract computation.

An analogy can be made to color. Colors can be coded by specific names, such as periwinkle, celandine and azure. Alternatively, one can code colors by particular combinations of three simple numbers that represent the amount of red, green and blue that make up that color. In the latter scheme, a color cell performing a projection onto the red axis would fire electrical impulses, or spikes, proportional to the amount of red in any color. Such a cell would fire at the same intensity for a brown or yellow color containing the same amount of red mixed in with other colors. Face cells use the same scheme, but instead of just three axes, there are 50. And instead of each axis coding the amount of red, green or blue, each axis codes the amount of deviation of the shape or appearance of any given face from an average face.

It would seem then that the Jennifer Aniston cells do not exist, at least not in the IT cortex. But single neurons responding selectively to specific familiar individuals may still be at work in a part of the brain that processes the output of face cells. Memory storage regions—the hippocampus and surrounding areas—may contain cells that help a person recognize someone from past experience, akin to the famed grandmother cells.

Facial recognition in the IT cortex thus rests on a set of about 50 numbers in total that represent the measurement of a face along a set of axes. And the discovery of this extremely simple code for face identity has major implications for our understanding of visual object representation. It is possible that all of the IT cortex might be organized along the same principles governing the face-patch system, with clusters of neurons encoding different sets of axes to represent an object. We are now conducting experiments to test this idea.

NEURAL ROSETTA STONE

IF YOU EVER GO to the British Museum, you will see an amazing artifact, the Rosetta stone, on which the same decree of Memphis is engraved in three different languages: Egyptian hieroglyphics, Demotic and ancient Greek. Because philologists knew ancient Greek, they could use the Rosetta stone to help decipher Egyptian hieroglyphics and Demotic. Similarly, faces, face patches and the IT cortex form a neural Rosetta stone—one that is still being deciphered. By showing pictures of faces to monkeys, we discovered face patches and learned how cells within these patches detect and identify faces. In turn, understanding coding principles in the face-patch network may one day lead to insight into the organization of the entire IT cortex, revealing the secret to how object identity more generally is encoded. Perhaps the IT cortex contains additional networks specialized for processing other types of objects—a whirring factory with multiple assembly lines.

UNDERSTANDING CODING PRINCIPLES IN THE FACE-PATCH NETWORK MAY ONE DAY LEAD TO INSIGHT INTO THE ORGANIZATION OF THE ENTIRE INFERO-TEMPORAL CORTEX, REVEALING THE SECRET TO HOW OBJECT IDENTITY MORE GENERALLY IS ENCODED.

I also hope that knowing the code for facial identity can help fulfill my college dream of discovering how we imagine curves. Now that we understand face patches, we can begin to train animals to imagine faces and explore how neural activity is shaped by the purely internal act of imagination. Lots of new questions arise. Does imagination reactivate the code for the imagined face in the face patches? Does it bring back even earlier representations of contours and shading that provide inputs to the face-patch system? We now have the tools to probe these questions and better understand how the brain sees objects, imagined or real.

Because almost all the brain's core behaviors—consciousness, visual memory, decision-making, language—require object interactions, a deep understanding of object perception will help us gain insight into the entire brain, not just the visual cortex. We are only starting to solve the enigma of the face. SA

Doris Y. Tsao is a professor of biology at the California Institute of Technology and an investigator of the Howard Hughes Medical Institute. In 2018 she was named a MacArthur Fellow.

RADICAL CHANGE

Uncertainty in the world threatens our sense of self. To cope, people embrace populism

By Michael A. Hogg

Illustration by Wesley Allsbrook

HUMAN SOCIETIES ARE CONSTANTLY REARRANGING THEMSELVES, CAUSING PROFOUND disruptions in our social lives. The industrial revolution of the late 18th and early 19th centuries fragmented communities as people moved for work, the decay of empires in the early 20th century reconfigured nations and national identities, and the Great Depression of the 1930s shattered people's economic security and future prospects. But we are now in what is perhaps a time of unprecedented uncertainty. The early 21st century is characterized by rapid and overwhelming change: globalization, immigration, technological revolution, unlimited access to information, sociopolitical volatility, the automation of work and a warming climate.

PDF downloaded @MOBILISM.ORG

Social groups soothe individual self-uncertainty but also delineate "in-groups" and "out-groups," which can increase populism.

People need to have a firm sense of identity and of their place in the world, and for many the pace and magnitude of such change can be alienating. This is because our sense of self is a fundamental organizing principle for our own perceptions, feelings, attitudes and actions. Typically it is anchored in our close interpersonal relationships, such as with our friends, family and partners, and in the variety of social groups and categories that we belong to and identify with—our nationality, religion, ethnicity, profession. It allows us to predict with some confidence how others will view us and treat us.

Imagine navigating all the situations and people we encounter in day-to-day life while continually feeling uncertain about who we are, how to behave and how social interactions will unfold. We would feel disorientated, anxious, stressed, cognitively depleted, and lacking agency and control. This self-uncertainty can, in fact, be experienced as an exciting challenge if we feel we have the material, social and psychological resources to resolve it. If we feel we do not have these resources, however, it can be experienced as a highly aversive threat to us and our place in the world.

Generally, self-uncertainty is a sensation that people are motivated to reduce. When people are increasingly unsure about who they are and how they fit into this rapidly changing landscape, it can be—and indeed has become—a real problem for society. People are supporting and enabling authoritarian leaders, flocking to ideologies and worldviews that promote and celebrate the myth of a glorious past. Fearful of those who are different from them, they seek homogeneity and become intoxicated by the freedom to access only information that confirms who they are or who they would like to be. As a result, global populism is on the rise.

SEEKING SOCIAL IDENTITY

ONE POWERFUL SOURCE of identity resides in social groups. They can be highly effective at reducing a person's self-uncertainty—particularly if such groups are distinctive and have members who share a sense of interdependence.

Groups play this central role in anchoring who we are because they are social categories, and research shows that social categorization is ubiquitous. A person categorizes others as either "in-group" or "out-group" members. They assign the group's attributes and social standing to those others, thereby constructing a subjective world where groups are internally homogeneous and the differences between groups are exaggerated and polarized in an ethnocentric manner. And because we also categorize ourselves, we internalize shared in-group-defining attributes as part of who we are. To build social identity, we psychologically surround ourselves with those who are like us.

This psychological process that causes people to identify with groups and behave as group members is called social categorization. It anchors and crystallizes our sense of self by assigning us an identity that prescribes how we should behave, what we should think and how we should make sense of the world. It also makes interaction predictable, allows us to anticipate how people will treat and think about us, and furnishes consensual identity confirmation: people like us—the in-group members—validate who we are.

This self-uncertainty social-identity dynamic is not in itself a bad thing. It enables the collective organization that lies at the heart of human society. Human achievements that require the coordination of many in the service of common goals cannot be achieved by individuals on their own. Yet this dynamic becomes a problem when the sense of self-uncertainty and identity threat is acute, enduring and all-encompassing. People then experience an overwhelming need for identity—and not just any identities but ones that are well equipped to resolve those disorienting, even scary, feelings.

REDUCING UNCERTAINTY THROUGH GROUP MEMBERSHIP

SOME FEATURES of groups and social identities are especially well suited to reducing self-uncertainty. Most important, groups need to be polarized from other groups and have unambiguous boundaries that

Pyrosky/Getty Images

distinguish between those who are "in" and those who are "out." Internally they need to be clearly structured, typically in a hierarchical way. These features make the group cohesive and homogeneous, such that members are interdependent and of one mind in sharing a common fate.

Diversity and dissent reinstate uncertainty and are therefore avoided. When these facets do occur, individuals and the group as a whole react decisively and harshly, creating an atmosphere of suspicion that lays the ground for persecution of alleged deviants. It breeds an opportunity for personal dislikes and vendettas to escalate under the guise of protecting cohesion.

That members are accepted and trusted fully is important not only for the group but also for the members themselves. After all, they desperately want to be included so that their identity is validated and their uncertainty thus reduced. Prospective and new members—and those who suspect they are viewed with suspicion or are uncertain about whether they are fully accepted—will go to extremes on behalf of the group to prove their membership credentials and loyalty. These individuals are vulnerable to zealotry and radicalization. Neo-Nazis and white supremacists who publicly engage in violent acts of terrorism and racial hatred are one example of this extremism.

The social identity embodied by such groups also needs to be uncomplicated so that it can be taken at face value as "the truth." Subtlety and nuance are anathema because they are an impediment to uncertainty reduction. Clarity on where the group stands allows its members to know how they should think and feel—as well as behave. Such identities are bolstered by having a strong ideology that identifies distasteful and morally bankrupt out-groups who can be demonized and cast in the role of "enemy." Conspiracy theories thrive in this environment because they establish these out-groups as agents of historical victimization by the in-group.

HOW UNCERTAINTY BREEDS POPULISM

IF SELF-UNCERTAINTY motivates people to identify with a group and internalize that identity as a key part of who they are, they need to be confident that they know exactly what their group's identity is. When you need what *you* consider to be reliable and trusted sources of identity information, where do you turn? The first port of call is whoever you believe is consensually viewed by the group as its leadership—typically someone whose leadership position is also formalized.

Recent research on how self-uncertainty affects the type of leaders that individuals prefer paints a potentially alarming picture. People just need someone to tell them what to do—and ideally those directives are coming from someone whom they can trust as "one of us." Self-uncertain people have also been shown to prefer leaders who are assertive and authoritarian, even autocratic, and who deliver a simple, black-and-white, affirmational message about "who we are" rather than a more open, nuanced and textured identity message.

Perhaps most troubling is that self-uncertainty can enable and build support for leaders who possess the so-called Dark Triad personality attributes: Machiavellianism, narcissism and psychopathy. Self-uncertainty, in other words, seems to fuel populism.

Another source of identity information is "people like you" who you feel embody the group's identity and see the world in the same way as you do. These can be people with whom you interact face-to-face or as friends, or they can be sources of information such as radio and television channels, particularly news outlets, that you watch. But nowadays these sources are overwhelmingly information and influence nodes on the Internet, such as Web sites, social media, Twitter feeds, podcasts, and so forth.

The Web is an ideal place to decrease the discomfort of self-uncertainty because it provides nonstop access to unlimited information that is often cherry-picked by individuals themselves and algorithms that do it discreetly. Therefore, people are accessing only identity-confirming information. Confirmation bias, a powerful and universal human bias that is especially strong under uncertainty, separates information and identity universes that fragment and polarize society. Online, people can easily seek out groups that may not be readily available in their physical lives.

WHEN PEOPLE ARE INCREASINGLY UNSURE ABOUT WHO THEY ARE AND HOW THEY FIT INTO THIS RAPIDLY CHANGING LANDSCAPE, IT CAN BE—AND INDEED HAS BECOME—A REAL PROBLEM FOR SOCIETY.

The Internet further empowers confirmation bias under uncertainty because people want to be surrounded by those who think alike so that their identities and worldview are continuously confirmed. The contours of "truth" then get mapped onto these self-contained social-identity universes. In this scenario, there are no absolute truths and no motivation to seriously explore and incorporate alternative viewpoints because that would be kryptonite to social identity's power to reduce self-uncertainty. This dynamic helps to explain why people dwell in increasingly homogeneous echo chambers that confirm their identity. SA

Michael A. Hogg is a professor and chair of social psychology at Claremont Graduate University and an honorary professor at the University of Kent in England. He is a former president of the Society of Experimental Social Psychology, an editor in chief of the journal *Group Processes & Intergroup Relations*, and a fellow of numerous societies, including the Association for Psychological Science.

Unconscious Mind

The Intention Machine *page 74*

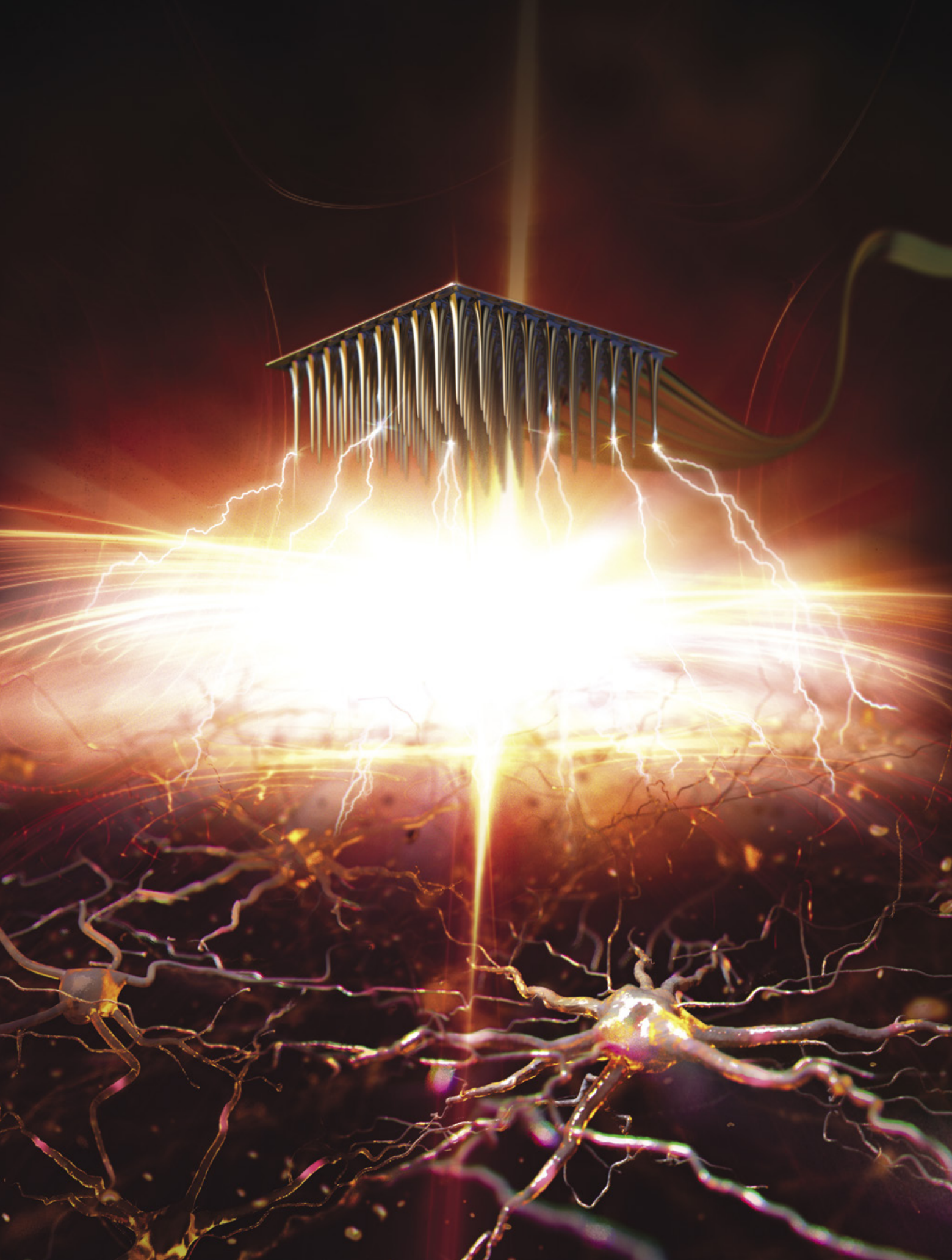
Infectious Dreams *page 82*

Sleep Learning Gets Real *page 88*

Answering Queries in Real Time while Dreaming *page 94*

Can Science Illuminate Our Inner Dark Matter? *page 96*

Illustration by Maria Corte



THE INTENTION MACHINE

A new generation of brain-machine interfaces can deduce what a person wants

By Richard A. Andersen

Illustration by Mark Ross

I GET GOOSE BUMPS EVERY TIME I SEE IT. A PARALYZED VOLUNTEER SITS IN a wheelchair while controlling a computer or robotic limb just with his or her thoughts—a demonstration of a brain-machine interface (BMI) in action.

That happened in my laboratory in 2013, when Erik Sorto, a victim of a gunshot wound when he was 21 years old, used his thoughts alone to drink a beer without help for the first time in more than 10 years. The BMI sent a neural message from a high-level cortical area. An electromechanical appendage was then able to reach out and grasp the bottle, raising it to Sorto's lips before he took a sip. His drink came a year after surgery to implant electrodes in his brain to control signals that govern the thoughts that trigger motor movement. My lab colleagues and I watched in wonderment as he completed this deceptively simple task that is, in reality, intricately complex.

Witnessing such a feat immediately raises the question of how mere thoughts can control a mechanical prosthesis. We move our limbs unthinkingly every day—and completing these motions with ease is the goal of any sophisticated BMI. Neuroscientists, though, have tried for decades to decode neural signals that initiate movements to reach out and grab objects. Limited success in reading these signals has spurred a search for new ways to tap into the cacophony of electrical activity resonating as the brain's 86 billion neurons communicate. A new generation of BMIs now holds the promise of creating a seamless tie between brain and prosthesis by tapping with great precision into the neural regions that formulate actions—whether the desired goal is grasping a cup or taking a step.

FROM BRAIN TO ROBOT

A BMI OPERATES by sending and receiving—"writing" and "reading"—messages to and from the brain. There are two major classes of the interface technology. A "write-in" BMI generally uses electrical stimulation to transmit a signal to neural tissue. Successful clinical applications of this technology are already in use. The cochlear prosthesis stimulates the auditory nerve to enable deaf subjects to hear. Deep-brain stimulation of an area that controls motor activity, the basal ganglia, treats motor disorders such as Parkinson's disease and essential tremor. Devices that stimulate the retina are currently in clinical trials to alleviate certain forms of blindness.

"Read-out" BMIs, in contrast, record neural activity and are still at a developmental stage. The unique challenges of reading neural signals need to be addressed before this next-generation technology reaches patients. Coarse read-out techniques already exist. The electroencephalogram (EEG) records the average activity over centimeters of brain tissue, capturing the activity of many millions of neurons rather than that from individual neurons in a single circuit. Functional magnetic resonance imaging (fMRI) is an indirect measurement that records an increase in blood flow to an active region. It can image smaller areas than EEG, but its resolution is still rather low. Changes in blood flow are slow, so fMRI cannot distinguish rapid changes in brain activity.

To overcome these limitations, ideally one would like to record the activity of individual neurons. Observing changes in the firing rate of large numbers of single neurons can provide the most complete picture of what is happening in a specific brain region. In recent years arrays of tiny electrodes implanted in the brain have begun to make this type of recording possible. The arrays now in use are four-by-four-millimeter flat surfaces with 100 electrodes. Each electrode, measuring one to 1.5 millimeters long, sticks out of the flat surface. The entire array, which resembles a bed of nails, can record activity from 100 to 200 neurons.

The signals recorded by these electrodes move to "decoders" that use mathematical algorithms to translate varied patterns of single-neuron firing into a signal that initiates a particular movement, such as control of a robotic limb or a computer. These read-out BMIs will assist patients who have sustained brain injury because of spinal cord lesions, stroke, multiple sclerosis, amyotrophic lateral sclerosis and Duchenne muscular dystrophy.

Our lab has concentrated on people with tetraplegia, who are unable to move either their upper or lower limbs because of upper spinal cord injuries. We make recordings from the cerebral cortex, the approximately three-millimeter-thick surface of

the brain's two large hemispheres. If spread flat, the cortex of each hemisphere would measure about 80,000 square millimeters. The number of cortical regions that specialize in controlling specific brain functions has grown as more data have been collected and is now estimated to encompass more than 180 areas. These locations process sensory information, communicate to other brain regions involved with cognition, make decisions or send commands to trigger an action.

In short, a brain-machine interface can interact with many areas of the cortex. Among them are the primary cortical areas, which detect sensory inputs, such as the angle and intensity of light impinging on the retina or the sensation triggered in a peripheral nerve ending. Also targeted are the densely connected association cortices between the primary areas that are specialized for language, object recognition, emotion and executive control of decision-making.

A handful of groups have begun to record populations of single neurons in people who are paralyzed, allowing them to operate a prosthesis in the controlled setting of a lab. Major hurdles still persist before a patient can be outfitted with a neural prosthetic device as easily as a heart pacemaker. My group is pursuing recordings from the association areas instead of the motor cortex targeted by other labs. Doing so, we hope, may provide greater speed and versatility in sensing the firing of neural signals that convey one's intentions.

The specific association area my lab has studied is the posterior parietal cortex (PPC), where plans to initiate movements begin. In our work with nonhuman primates, we found one subarea of the PPC, called the lateral intraparietal cortex, that discerns intentions to begin eye movements. Limb-movement processing occurs elsewhere in the PPC. The parietal reach region prepares arm movements. Also, Hideo Sakata, then at the Nihon University School of Medicine in Japan, and his colleagues found that the anterior intraparietal area formulates grasping movements.

Recordings from nonhuman primates indicate that the PPC provides several possible advantages for brain control of robotics or a computer cursor. It controls both arms, whereas the motor cortex in each hemisphere, the area targeted by other labs, activates primarily the limb on the opposite side of the body. The PPC also indicates the goal of a movement. When a nonhuman primate, for instance, is visually cued to reach for an object, this brain area switches on immediately, flagging the location of a desired object. In contrast, the motor cortex sends a signal for the path the reaching movement should take. Knowing the goal of an intended motor action lets the BMI decode it quickly, within a couple of hundred milliseconds, whereas figuring out the trajectory signal from the motor cortex can take more than a second.

INTERFACE TECHNOLOGY, developed by Richard A. Andersen (*left*) and his Caltech team, enabled Erik Sorto (*right*) to move a robotic arm.

Lance Hayashida and California Institute of Technology

FROM LAB TO PATIENT

IT WAS NOT EASY to go from experiments in lab animals to studies of the PPC in humans. Fifteen years elapsed before we made the first human implant. First, we inserted the same electrode arrays we planned to use in humans into healthy nonhuman primates. The monkeys then learned to control computer cursors or robotic limbs.

We built a team of scientists, clinicians and rehabilitation professionals from the California Institute of Technology, the University of Southern California, the University of California, Los Angeles, the Rancho Los Amigos National Rehabilitation Center, and Casa Colina Hospital and Centers for Healthcare. The team received a go-ahead from the Food and Drug Administration and institutional review boards charged with judging the safety and ethics of the procedure in the labs, hospitals and rehabilitation clinics involved.

A volunteer in this type of project is a true pioneer because he or she may or may not benefit. Participants ultimately join

to help users of the technology who will seek it out once it is perfected for everyday use. The implant surgery for Sorto, our first volunteer, took place in April 2013 and was performed by neurosurgeons Charles Liu and Brian Lee. The procedure went flawlessly, but then came the wait for healing before we could test the device.

My colleagues at NASA's Jet Propulsion Laboratory, which built and launched the Mars rovers, talk about the seven minutes of terror when a rover enters the planet's atmosphere before it lands. For me it was two weeks of trepidation, wondering whether the implant would work. We knew in nonhuman primates how similar areas of the brain functioned, but a human implant was testing uncharted waters. No one had ever tried to record from a population of human PPC neurons before.

During the first day of testing we detected neural activity, and by the end of the week there were signals from enough neurons to begin to determine if Sorto could control a robot limb. Some of the neurons varied their activity when Sorto imagined rotating his hand. His first task consisted of turning the robot hand to different orientations to shake hands with a graduate student. He was thrilled, as were we, because this accomplishment marked the first time since his injury he could interact with the world using the bodily movement of a robotic arm.

People often ask how long it takes to learn to use a BMI. In fact, the technology worked right out of the box. It was intuitive and easy to use the brain's intention signals to control the robotic arm. By imagining different actions, Sorto could watch recordings of individual neurons from his cortex and turn them on and off at will.

We ask participants at the beginning of a study what they would like to achieve by controlling a robot. For Sorto, he wanted to be able to drink a beer on his own rather than asking someone else for help. He was able to master this feat about one year into the study. With the team co-led by research scientist Spencer Kellis of Caltech, which included roboticists from the Applied Physics Laboratory at Johns Hopkins University, we melded Sorto's intention signals with the processing power furnished by machine vision and smart robotic technology.

The vision algorithm analyzes inputs from video cameras, and the smart robot combines the intent signal with computer algorithms to initiate the movement of the robot arm. Sorto achieved this goal after a year's time with cheers and shouts of joy from everyone present. In 2015 we published in *Science* our first results on using intention signals from the PPC to control neural prostheses.

Sorto is not the only user of our technology. Nancy Smith, now in her fourth year in the study, became tetraplegic from an automobile accident about 10 years ago. She had been a high school teacher of computer graphics and played piano as a pastime. In our studies with lead team members Tyson Aflalo of Caltech and Nader Pouratian of U.C.L.A., we found a detailed representation of the individual digits of both hands in Smith's PPC. Using virtual reality, she could imagine and move 10 fingers individually on left and right "avatar" hands displayed on a computer screen. Using the imagined movement of five fingers from one hand, Smith could play simple melodies on a computer-generated piano keyboard.

By Thought Alone

For 15 years neuroscientists have built brain-machine interfaces (BMIs) that allow neural signals to move computer cursors or operate prostheses. The technology has moved forward slowly because translating the electrical firing of neurons into commands to play a video game or move a robot arm involves highly intricate processes.

A group at the California Institute of Technology has tried to advance the neuroprosthetic field by tapping into high-level neural processing—the intent to initiate an action—and then conveying the relevant electrical signals to a robotic arm. Instead of sending out signals from the motor cortex to move an arm, as attempted by other laboratories, the Caltech researchers place electrodes in the posterior parietal cortex (PPC), which transmits to a prosthesis the brain's intent to act.

Decoding neural signals remains a challenge for neuroscientists. But using BMI signals from the posterior parietal cortex, the top of the cognitive command chain, appears to result in faster, more versatile control of prosthetic technology.

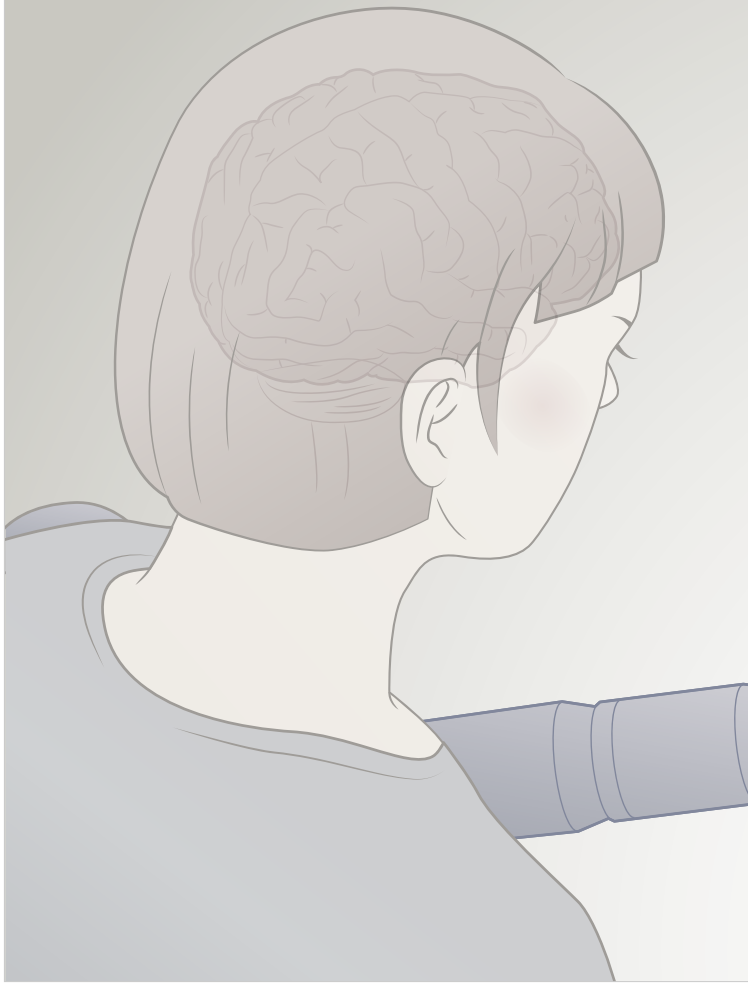
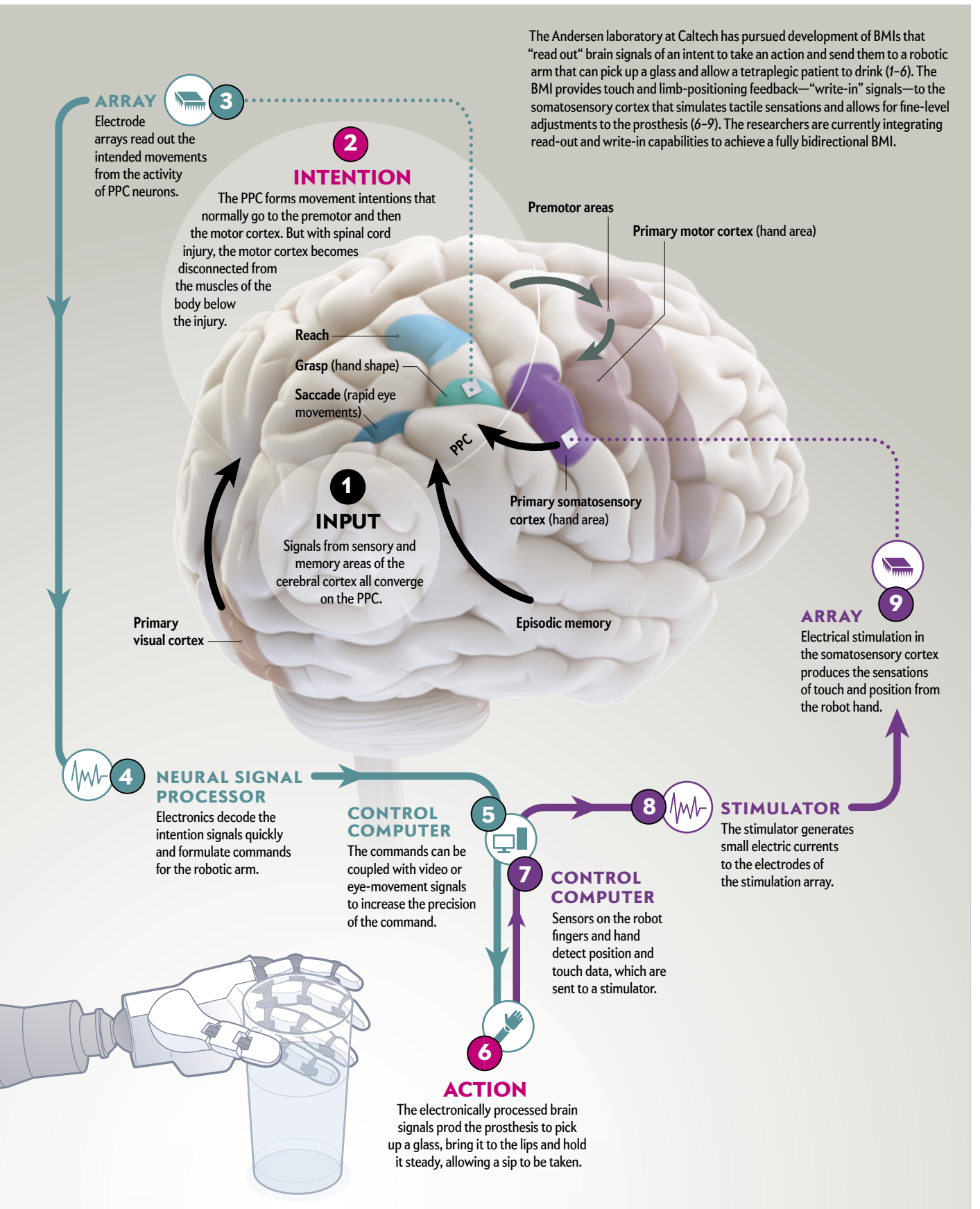


Illustration by AXS Biomedical Animation Studio

The Andersen laboratory at Caltech has pursued development of BMIs that "read out" brain signals of an intent to take an action and send them to a robotic arm that can pick up a glass and allow a tetraplegic patient to drink (*1–6*). The BMI provides touch and limb-positioning feedback—"write-in" signals—to the somatosensory cortex that simulates tactile sensations and allows for fine-level adjustments to the prosthesis (*6–9*). The researchers are currently integrating read-out and write-in capabilities to achieve a fully bidirectional BMI.


3
ARRAY
Electrode arrays read out the intended movements from the activity of PPC neurons.
2
INTENTION
The PPC forms movement intentions that normally go to the premotor and then the motor cortex. But with spinal cord injury, the motor cortex becomes disconnected from the muscles of the body below the injury.
Premotor areas
Primary motor cortex (hand area)
Reach
Grasp (hand shape)
Saccade (rapid eye movements)
PPC
1
INPUT
Signals from sensory and memory areas of the cerebral cortex all converge on the PPC.
Primary somatosensory cortex (hand area)
Primary visual cortex
Episodic memory
9
ARRAY
Electrical stimulation in the somatosensory cortex produces the sensations of touch and position from the robot hand.
4
NEURAL SIGNAL PROCESSOR
Electronics decode the intention signals quickly and formulate commands for the robotic arm.
CONTROL COMPUTER
The commands can be coupled with video or eye-movement signals to increase the precision of the command.
5
8
STIMULATOR
The stimulator generates small electric currents to the electrodes of the stimulation array.
7
CONTROL COMPUTER
Sensors on the robot fingers and hand detect position and touch data, which are sent to a stimulator.
6
ACTION
The electronically processed brain signals prod the prosthesis to pick up a glass, bring it to the lips and hold it steady, allowing a sip to be taken.

HOW THE BRAIN REPRESENTS GOALS

WE WERE THRILLED in working with these patients to find neurons tuned to processing signals related to one's intentions. The amount of information to be gleaned from just a few hundred neurons turned out to be overwhelming. We could decode a range of cognitive activity, including mental strategizing (imagined versus attempted motion), finger movements, decisions about recalling visual stimuli, hand postures for grasping, observed actions, action verbs such as "grasp" or "push," and visual and somatosensory perception. To our surprise, inserting a few tiny electrode arrays enabled us to decode much of what a person intends to do, as well as the sensory inputs that lead to the formation of intentions.

The question of how much information can be recorded from a small patch of brain tissue reminded me of a similar scientific problem that I had encountered early in my career. During my postdoctoral training with the late Vernon Mountcastle at the Johns Hopkins University School of Medicine, we examined how visual space is represented in the PPC of monkeys. Our eyes are like cameras, with the photosensitive retinas signaling the location of visual stimuli imaged on them—the entire image is referred to as a retinotopic map. Neurons respond to limited regions of the retina, referred to as their receptive fields. In other ways, processing visual perception is different than a video camera recording. When a video camera moves around, the recorded image also shifts, but when we move our eyes the world seems stable. The retinotopic image coming from the eyes must be converted into a visual representation of space that takes into account where the eyes are looking so that as they move, the world does not appear as if it were sliding around.

INSERTING A FEW TINY ELECTRODE ARRAYS INTO THE BRAIN ENABLED US TO DECODE MUCH OF WHAT A PERSON INTENDS TO DO.

The PPC is a key processing center for high-order visual space representation. For a person to reach and grab an object, the brain needs to take into account where the eyes are looking to pick it up. PPC lesions in humans produce inaccurate reaching. In Mountcastle's lab, we found individual PPC neurons had receptive fields that registered parts of a scene. The same cells also carried eye-position information. The two signals interacted by multiplying the visual response by the position of the eyes in the head—the product of which is called a gain field.

I continued to pursue this problem of understanding the brain's representation of space when I took my first faculty position at the Salk Institute for Biological Studies, right across the street from the University of California, San Diego. Working with David Zipser, a U.C.S.D. theoretical neuroscientist developing neural networks, we reported in *Nature* on a computational model of a neural network that combined retinotopic locations with gaze direction to make maps of space that are invariant to eye movements. During training of the neural networks, their middle layers developed gain fields, just as was the case in the PPC experiments. By mixing signals for visual inputs and eye positions within the same neurons, as few as nine neurons could represent the entire visual field.

Recently this idea of mixed representations—populations of neurons responding to multiple variables (as with the gain fields)—has attracted renewed attention. For instance, recordings from the prefrontal cortex show a mixing of two types of memory task and different visual objects.

This work, moreover, may have a direct bearing in explaining what is happening in the PPC. We discovered this when we asked Smith, using a set of written instructions, to perform eight different combinations of a task. One of her undertakings required strategizing to imagine or attempt an action. Another necessitated using the right and left side of the body; a third entailed squeezing a hand or shrugging a shoulder. We found that PPC neurons mixed all these variables—and the intermingling exhibited a specific pattern, unlike the random interactions we and others had reported in lab animal experiments.

Activity of populations of neurons for strategizing and for controlling each body side tends to overlap. If a neuron fires to initiate the movement of the left hand, it will most likely also respond for an attempted righthand movement, whereas neuron groups that control the shoulder and hand are more separated. We refer to this type of representation as partially mixed selectivity. We have since found similarities in partially mixed representations that seem to make up a semantics of movement. The activity of cells tuned for the same action type tends to overlap. A neuron that responds to videos of a person grasping an object will also likely become active when a person reads the word "grasp." But cells responding to an action such as pushing tend to get separated into their own group. In general, partially mixed coding appears to underlie computations that are similar (movements of the left hand are similar to those of the right). It also separates those that exhibit varying forms of neural processing (movement of the shoulder differs from movement of the hand).

Mixed and partially mixed coding have been found in certain parts of the association cortex—and new studies must explore whether they appear in other locations that govern language, object recognition and executive control. Additionally, we would like to know whether the primary sensory or motor cortical regions use a similar partially mixed structure.

Current studies indicate that, at least in the somatosensory cortex, neurons do not respond to visual stimuli or the intention to make a movement but do respond to somatosensory stimuli and to the imagined execution of movements. Thus, there is direct evidence that variables seen in the human PPC are not found in the primary somatosensory cortex, although it is still possible that partially mixed selectivity may exist in both areas but for different sets of variables.

Another near-future goal is to find out how much learning new tasks can affect the performance of the volunteers using the prosthesis. If learning readily takes place, any area of the brain might then be implanted and trained for any conceivable

BMI task. For instance, an implant in the primary visual cortex could learn to control motor tasks. But if learning is more restricted, an implant would be needed in a motor area to perform motor tasks. Early results suggest this latter possibility, and an implant may have to be placed in the area that has been previously identified as controlling particular neural functions.

WRITING IN SENSATIONS

A BMI MUST DO MORE than just receive and process brain signals—it must also send feedback from a prosthesis to the brain. When we reach to pick up an object, visual feedback helps to direct the hand to the target. The positioning of the hand depends on the shape of the object to be grasped. If the hand does not receive touch and limb-positioning signals once it begins to manipulate the object, performance degrades quickly.

Finding a way to correct this deficit is critical for our volunteers with spinal cord lesions, who cannot move their body below the injury. They also do not perceive the tactile sensations or positioning of their body that are essential to fluid movement. An ideal neural prosthesis, then, must compensate through bidirectional signaling: it must transmit the intentions of the volunteer but also detect the touch and positioning information arriving from sensors on a robotic limb.

Robert Gaunt and his colleagues at the University of Pittsburgh have addressed this issue by implanting microelectrode arrays in the somatosensory cortex of a tetraplegic person—where inputs from the limbs process feelings of touch. Gaunt's lab sent small electric currents through the microelectrodes, and the subject reported sensations from parts of the surface of the hand.

We have also used similar implants in the arm region of the somatosensory cortex. To our pleasant surprise, our subject, FG, reported natural sensations such as squeezing, tapping and vibrations on the skin, known as cutaneous sensations. He also perceived the feeling that the limb was moving—a sensation referred to as proprioception. These experiments show that subjects who have lost limb sensation can regain it through BMIs that have write-in perceptions. The next step is to provide a rich variety of somatosensory feedback sensations to improve robotic manual dexterity under brain control. Toward this goal, the Pittsburgh group has recently shown that stimulation of the primary somatosensory cortex improves the time to grasp objects with a robot limb, compared with standard visual feedback only. Also, we would like to know if subjects detect a sense of "embodiment," in which the robot limb appears to become part of their body.

As these clinical studies show us, both writing in and reading out cortical signals, provide insight into the degree of reorganization of the cerebral cortex after neurological injury. Numerous studies have reported a high degree of reorganization, but until recently there has been little focus on the fundamental structure that remains intact. BMI studies show that tetraplegic subjects can quickly use the motor and the PPC cortex to control assistive devices, and stimulation of the somatosensory cortex produces sensations in deinnervated areas that are similar to what would be expected for intact individuals. These results demonstrate considerable stability of the adult cortex even after severe injury and in spite of injury-induced plasticity.

FUTURE CHALLENGES

A MAJOR FUTURE CHALLENGE is to develop better electrodes for sending and receiving neural signals. We have found that current implants continue to function for a relatively lengthy five years. But better electrodes would ideally push the longevity of these systems even further and increase the number of neurons that can be recorded from them. Another priority—an increase in the lengths of the electrodes' tiny spikes—would help access areas located within folds of the cortex.

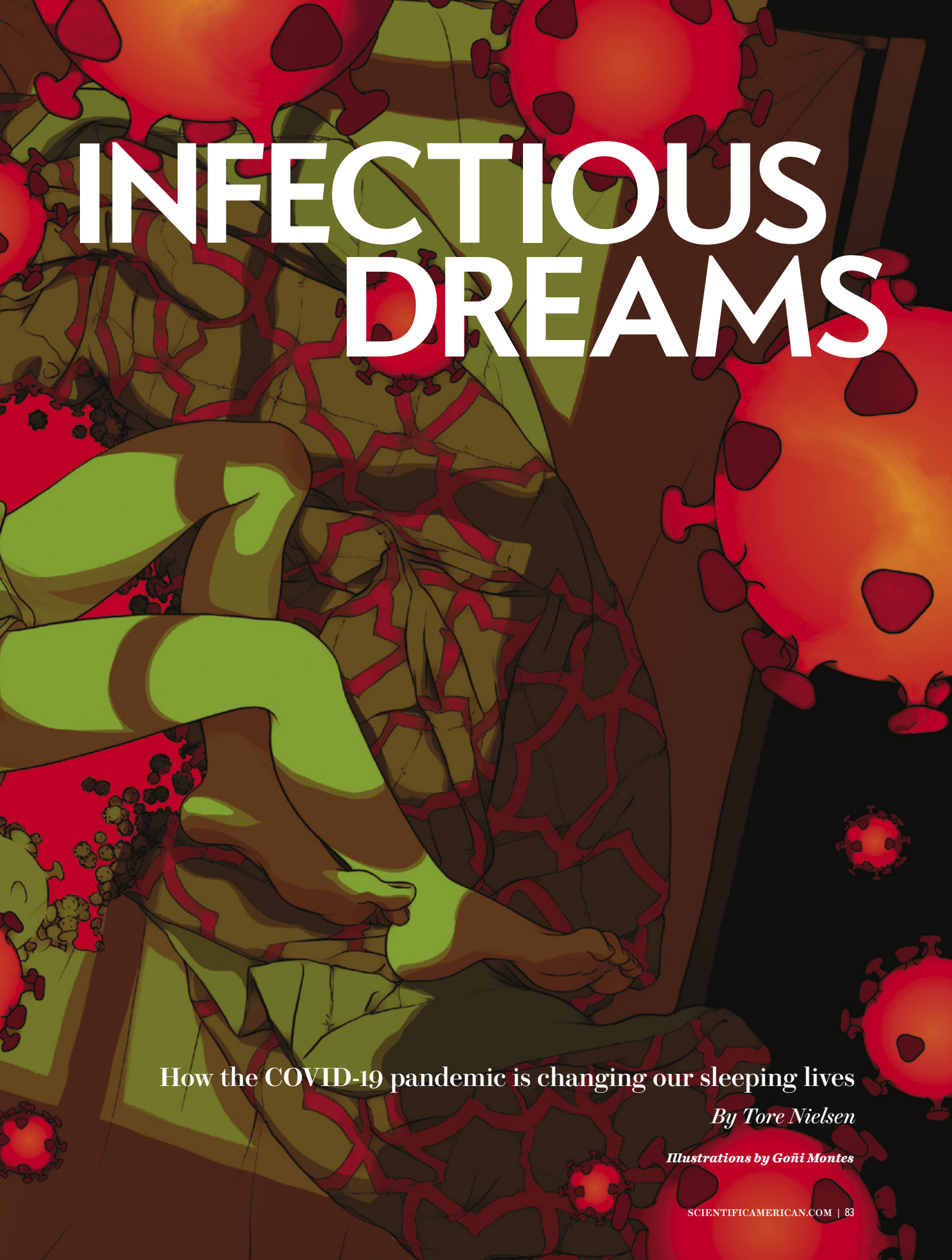
Flexible electrodes, which move with the slight jostling of the brain—from changes in blood pressure or the routine breathing cycle—will also allow for more stable recordings. Existing electrodes require recalibrating the decoder because the stiff electrodes change position with respect to neurons from day to day; researchers would ultimately like to follow the activity of identical neurons over weeks and months.

The implants need to be miniaturized, operate on low power (to avoid heating the brain), and function wirelessly so no cables are needed to connect the device to brain tissue. All current BMI technology needs to be implanted with a surgical procedure. But one day, we hope, recording and stimulation interfaces will be developed that can receive and send signals less invasively but with high precision. One step in this direction is our recent finding in nonhuman primates that ultrasound recorded changes in blood volume linked to neural activity can be used for BMIs. Because the skull is an impediment to ultrasound, a small ultrasound-transparent window would still be needed to replace a bit of the skull, but this surgery would be far less invasive than implanting microelectrode arrays that require opening the dura mater, the strong layer surrounding and protecting the brain, and directly inserting electrodes into the cortex.

BMIs, of course, are aimed at assisting people with paralysis. Yet science-fiction books, movies and the media have focused on the use of the technology for enhancement, conferring "superhuman" abilities that might allow a person to react faster, certainly an advantage for many motor tasks, or directly send and receive information from the cortex, much like having a small cell phone implanted in the brain. But enhancement is still very much in the realm of science fiction and will be achieved only when noninvasive technologies are developed that can operate at or near the precision of current microelectrode array technology.

Finally, I would like to convey the satisfaction of doing basic research and making it available to patients. Fundamental science is necessary to both advance knowledge and develop medical therapies. To be able to then transfer these discoveries into a clinical setting brings the research endeavor to its ultimate realization. A scientist is left with an undeniable feeling of personal fulfillment in sharing with patients their delight at being able to move a robotic limb to interact again with the physical world. SA

Richard A. Andersen is James G. Boswell Professor of Neuroscience and the Tianqiao and Chrissy Chen Brain-Machine Interface Center Leadership chair and center director at the California Institute of Technology. He studies the neural mechanisms of sight, hearing, balance, touch and action, as well as the development of neural prostheses. Andersen is a member of the National Academy of Sciences and the National Academy of Medicine.



INFECTIOUS DREAMS

How the COVID-19 pandemic is changing our sleeping lives

By Tore Nielsen

Illustrations by Goñi Montes

FOR MANY OF US, LIVING IN A COVID-19 WORLD FEELS AS IF WE HAVE BEEN thrown into an alternative reality. We live day and night inside the same walls. We fear touching groceries that arrive at our doorstep. If we venture into town, we wear masks, and we get anxious if we pass someone who is not wearing one. We have trouble discerning faces. It's like living in a dream.

COVID has altered our dream worlds, too: how much we dream, how many of our dreams we remember and the nature of our dreams themselves. In early 2020, when stay-at-home directives were put in place widely, society quite unexpectedly experienced what I am calling a dream surge: a global increase in the reporting of vivid, bizarre dreams, many of which are concerned with coronavirus and social distancing. Terms such as coronavirus dreams, lockdown dreams and COVID nightmares emerged on social media. By April of that year, social and mainstream media outlets had begun broadcasting the message: the world is dreaming about COVID.

Although widespread changes in dreaming had been reported in the U.S. following extraordinary events such as the 9/11 attacks in 2001 and the 1989 San Francisco earthquake, a surge of this magnitude had never been documented. This upwelling of dreams is the first to occur globally and the first to happen in the era of social media, which makes dreams readily accessible for immediate study. As a dream "event," the pandemic is unprecedented.

But what kind of phenomenon is this, exactly? Why was it happening with such vigor? To find out, Deirdre Barrett, an assistant professor at Harvard Medical School and editor in chief of the journal *Dreaming,* initiated a COVID dreams survey online in the week of March 22, 2020. Erin and Grace Gravley, San Francisco Bay Area artists, launched IDreamofCovid.com, a site archiving and illustrating pandemic dreams. The Twitter account @CovidDreams began operation. Kelly Bulkeley, a psychologist of religion and director of the Sleep and Dream Database, followed with a YouGov survey of 2,477 American adults. And my former doctoral student Elizaveta Solomonova, now a postdoctoral fellow at McGill University, along with Rébecca Robillard of the Royal's Institute of Mental Health Research in Ottawa and others, launched a survey to which 968 people aged 12 and older responded, almost all in North America. Results of these inquiries, published in *BMJ Open* in December 2020, document the precipitous surge, the striking variety of dreams and many related mental health effects.

Bulkeley's three-day poll revealed that in March 2020, 29 percent of Americans recalled more dreams than usual. Solomonova and Robillard found that 37 percent of people had pandemic dreams, many marked by themes of insufficiently completing tasks (such as losing control of a vehicle) and being threatened by others. Many online posts from the time reflect these findings. One person, whose Twitter handle is @monicaluhar, reported, "*Had a dream about returning as a sub teacher in the fall, unprepared. Students were having a difficult time practicing social distancing, and teachers couldn't stagger classes or have one-on-one meetings.*" And @therealbeecarey said, "*My phone had a virus and was posting so many random pictures from my camera roll to instagram and my anxiety was at an all time high.*"

More recent studies found qualitative changes in dream emotions and concerns about health. Dream reports from Brazilian adults in social isolation had high proportions of words related to anger, sadness, contamination and cleanliness. Text

mining of accounts of 810 Finnish dreams showed that most word clusters were laden with anxiousness; 55 percent were about the pandemic directly (lack of regard for social distancing, elderly people in trouble), and these emotions were more prevalent among people who felt increased stress during the day. A study of 100 nurses conscripted to treat COVID patients in Wuhan, China, revealed that 45 percent experienced nightmares—twice the lifetime rate among Chinese psychiatric outpatients and many times higher than that among the 5 percent or so of the general population who have nightmare disorder.

It seems clear that some basic biological and social dynamics may have played a role in this unprecedented opening of the oneiric floodgates. At least three factors may have triggered or sustained the dream surge: disrupted sleep schedules augmenting the amount of rapid eye movement (REM) sleep and therefore dreaming; threats of contagion and social distancing taxing dreaming's capacity to regulate emotions; and social and mainstream media amplifying the public's reaction to the surge.

MORE REM SLEEP, MORE DREAMS

ONE OBVIOUS EXPLANATION for the surge is that sleep patterns changed abruptly when lockdowns took effect. Early publications demonstrate elevated levels of insomnia in China's population, especially among frontline workers. In contrast, stay-at-home orders, which removed long commutes to work, improved sleep for many people. Respondents in China reported an average increase of 46 minutes in bed and an extra 34 minutes in total sleep time. Some 54 percent of people in Finland said they slept more after lockdown. Overall, from March 13 to 27, 2020, time asleep in the U.S. increased almost 20 percent nationwide, and states with the longest commute times, such as Maryland and New Jersey, showed the largest increases.

Longer slumber leads to more dreams; people in sleep laboratories who are allowed to snooze for more than 9.5 hours recall more dreams than when sleeping a typical eight hours. Sleeping longer also proportionally increases REM sleep, which is when the most vivid and emotional dreams occur.

Relaxed schedules may also have caused dreaming to occur later than usual in the morning, when REM sleep is more prevalent and intense and, thus, dreams are more bizarre. Dream tweets reflect these qualities: "*I was taking care of a newborn girl that had COVID… it was so vivid and real*." Increased dreaming during late-morning REM intervals results from the convergence of several processes. Sleep itself cycles through deep and light stages about every 90 minutes, but pressure for REM sleep gradually increases as the need for deep, recuperative sleep is progressively satisfied. Meanwhile a circadian process that is tightly linked to our 24-hour core body temperature rhythm gives an abrupt boost to REM sleep propensity late in the sleep period and stays elevated through the morning.

After the pandemic began, many people did sleep longer and later. In China, average weekly bedtime was delayed by 26 minutes but wake-up time by 72 minutes. These values were 41 and 73 minutes in Italy and 30 and 42 minutes among U.S. university students. And without commutes, many people were freer to linger in bed, remembering their dreams. Some early birds may have turned into night owls, who typically have more REM sleep and more frequent nightmares. And as people eliminated whatever sleep debts they may have accrued over days or even weeks of insufficient rest, they were more likely to wake up at night and remember more dreams.

DREAM FUNCTIONS OVERWHELMED

THE SUBJECT MATTER of many COVID dreams directly or metaphorically reflects fears about contagion and the challenges of social distancing. Even in normal times, we dream more about novel experiences. For instance, people enrolled in programs to rapidly learn French dream more about French. Replaying fragments of experiences is one example of a functional role that researchers widely ascribe to REM sleep and dreaming: it helps us solve problems. Other roles include consolidating the prior day's events into longer-lasting memories, fitting those events into an ongoing narrative of our lives and helping us regulate emotions.

Researchers have documented countless cases of dreams assisting in creative achievement. Empirical studies also show that REM sleep aids in problem-solving that requires access to wide-ranging memory associations, which may explain why so many dreams in the 2020 surge involve creative or strange attempts to deal with a COVID problem. One survey respondent said, "*I was looking for a kind of cream that would either prevent or cure Covid-19. I got my hands on the last bottle*."

Two other widely claimed dream functions are extinguishing fearful memories and simulating social situations. They are related to emotion regulation and help to explain why pandemic threats and social distancing challenges appear so often in surge dreams. Many dreams reported in the media include fearful reactions to infection, finances and social distancing: "*I tested positive for pregnancy and covid … now I'm stressed*." Threats may take the form of metaphoric imagery such as tsunamis or aliens; zombies are common. Images of insects, spiders and other small creatures are also widely represented: "*My foot was covered in ants and 5-6 black widows were imbedded in the bottom of my foot.*"

One way to understand direct and metaphoric imagery is to consider that dreams express an individual's core concerns, drawing on memories that are similar in emotional tone but different in subject matter. This contextualization is clear in post-traumatic nightmares, in which a person's reaction to a trauma, such as terror during an assault, is depicted as terror in the face of, for example, a natural disaster such as a tsunami. The late Ernest Hartmann, a Boston-area dream and nightmare research pioneer who studied dreams after the 9/11 attacks, stipulated that such contextualization best helps people adapt when it weaves together old and new experiences. Successful integration produces a more stable memory system that is resilient to future traumas.

Metaphorical images can be part of a constructive effort to make sense of disruptive events. A related process is the extinguishing of fear by the creation of new "safety memories." These possibilities, which I and others have investigated, reflect the fact that memories of fearful events are almost never replayed in their entirety during dreaming. Instead elements of a memory appear piecemeal, as if the original memory has been reduced to basic units. These elements recombine with newer memories and cognitions to create contexts in which metaphors and other unusual juxtapositions of imagery seem incongruous or incompatible with waking life—and, more important, are incompatible with feelings of fear. This creative dreaming produces safety imagery that supersedes and inhibits the original fear memory, helping to assuage distress over time.

This mechanism can break down after severe trauma, however. When this happens, nightmares arise in which the fearful memory is replayed realistically; the creative recombining of memory elements is thwarted. The pandemic's ultimate impact on a person's dreams will vary with whether or how severely they are traumatized and how resilient they are.

A second class of theories—also still speculative—may explain social distancing themes, which permeated IDreamofCovid.com reports. Emotions in these dreams range from surprise to discomfort to stress to nightmarish horror. Tweets located by the @CovidDreams account illustrate how incompatible dream scenarios are with social distancing—so incompatible that they often trigger a rare moment of self-awareness and awakening: "*We were celebrating something by having a party. And I woke myself up because something wasn't right because we're social distancing and not supposed to be having parties.*"

These theories focus on dreaming's social simulation function. The view that dreaming is a neural simulation of reality, analogous to virtual reality, is now widely accepted, and the notion that the simulation of social life is an essential biological function is emerging. In 2000 Anne Germain, now CEO of sleep medicine start-up Noctem, and I proposed that images of characters interacting with the self in dreams could be basic to how dreaming evolved, reflecting attachment relationships essential to the survival of prehistoric groups. The strong interpersonal bonds reiterated during dreaming contribute to stronger group structures that help to organize defenses against predators and cooperation in problem-solving. Such dreams would still have adaptive value today because family and group cohesion remain essential to health and survival. It may be the case that an individual's concerns about other people are fine-tuned while they are in the simulated presence of those people. Important social relationships and conflicts are portrayed realistically during dreaming.

Other investigators, such as cognitive neuroscientist Antti Revonsuo of the University of Turku in Finland and the University of Skövde in Sweden, have since proposed additional social functions for dreaming: facilitating social perception (Who is around me?), social mind reading (What are they thinking?) and the practice of social bonding skills. Another theory advanced by psychology professor Mark Blagrove of Swansea University in Wales further postulates that by sharing dreams, people enhance empathy toward others. The range of dream functions is likely to keep expanding as we learn more about the brain circuits underlying social cognition and the roles REM sleep plays in memory for emotional stimuli, human faces and reactions to social exclusion. Because social distancing is, in effect, an experiment in social isolation at a level never before seen—and is likely antagonistic to human evolution—a clash with deep-rooted dream mechanisms should be evident on a massive scale. And because social distancing disrupts normal relationships so profoundly—causing many of us to spend excessive time with some people and no time with others—social simulations in dreams may play a crucial role in helping families, groups, even societies deal with sudden, widespread social adaptation.

THE ECHO CHAMBER OF SOCIAL MEDIA

THERE IS ONE BASIC QUESTION about pandemic dreams that we would like to nail down: whether the dream surge was amplified by the media. It is quite possible that early posts of a few dreams were circulated widely online, feeding a narrative of pandemic dreams that went viral, influencing people to recall their dreams, notice COVID themes and share them. This narrative may have even induced people to dream more about the pandemic.

Evidence suggests that mainstream media reporting probably did not trigger the surge but may have temporarily amplified its scope. The Bulkeley and Solomonova-Robillard polls corroborated a clear groundswell in dream tweeting during March 2020, before the first media stories about such dreams appeared; indeed, the earliest stories cited various tweet threads as their sources.

Once stories emerged, more surges in dream reporting through early April 2020 were detected by IDreamofCovid.com and @CovidDreams. The format of most early stories almost guaranteed amplification: they typically described some salient dream themes observed in a survey and provided a link directing readers to participate in the same survey. In addition, 56 percent of articles during the first week of stories featured interviews with the same Harvard dream scientist, which may have influenced readers to dream about themes repeated by her.

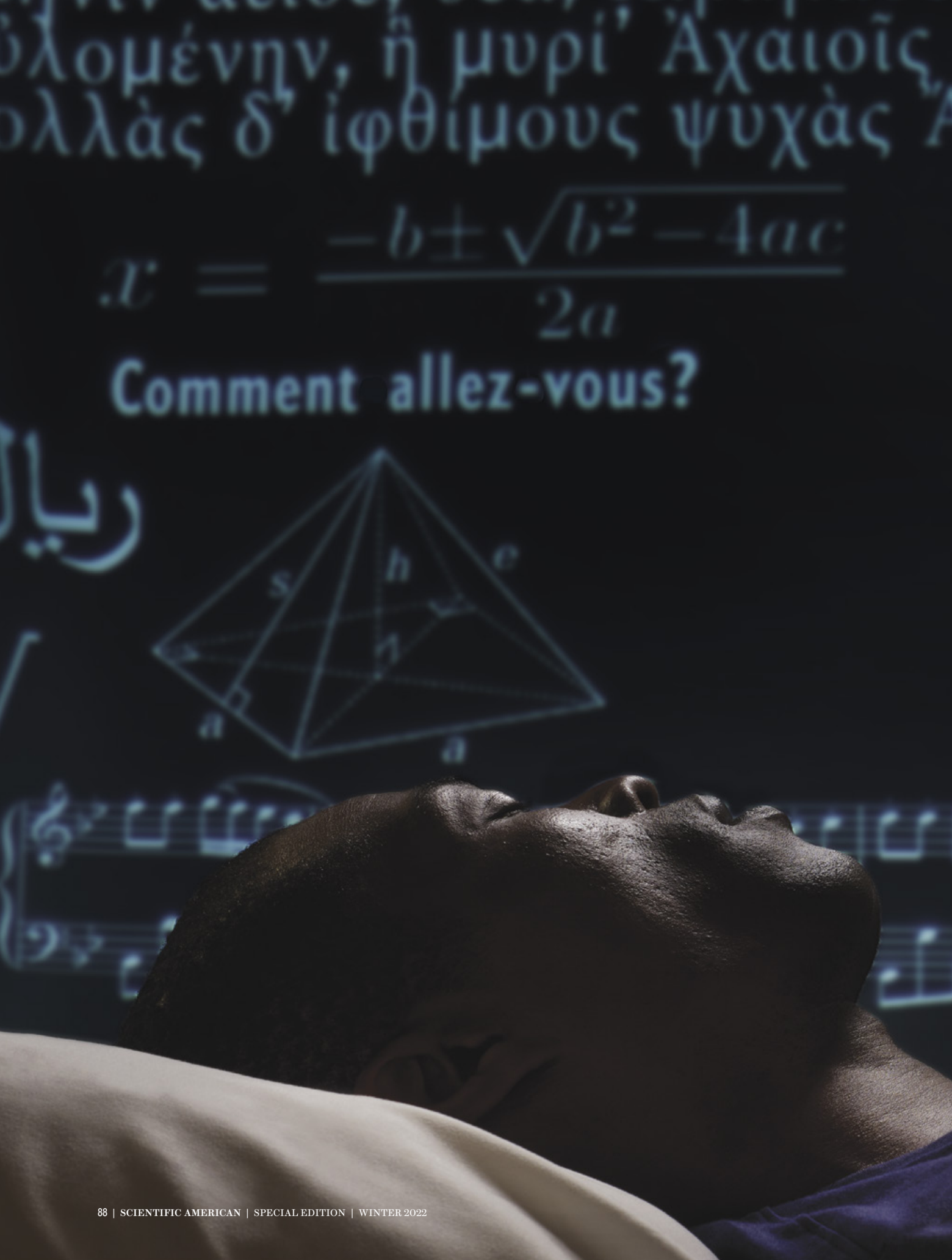
The surge began to decline steadily in late April 2020, as did the number of mainstream media articles, suggesting that any echo-chamber effect had run its course. The final nature of the surge remains to be seen. Until COVID vaccines or treatments are fully distributed and with waves of future infections or new viral variants possible, threats of disease and social distancing are likely to persist. Might the pandemic have produced a lasting increase in humanity's recall of dreams? Could pandemic concerns become permanently woven into dream content? And if so, will such alterations help or hinder people's long-term adjustments to our postpandemic futures?

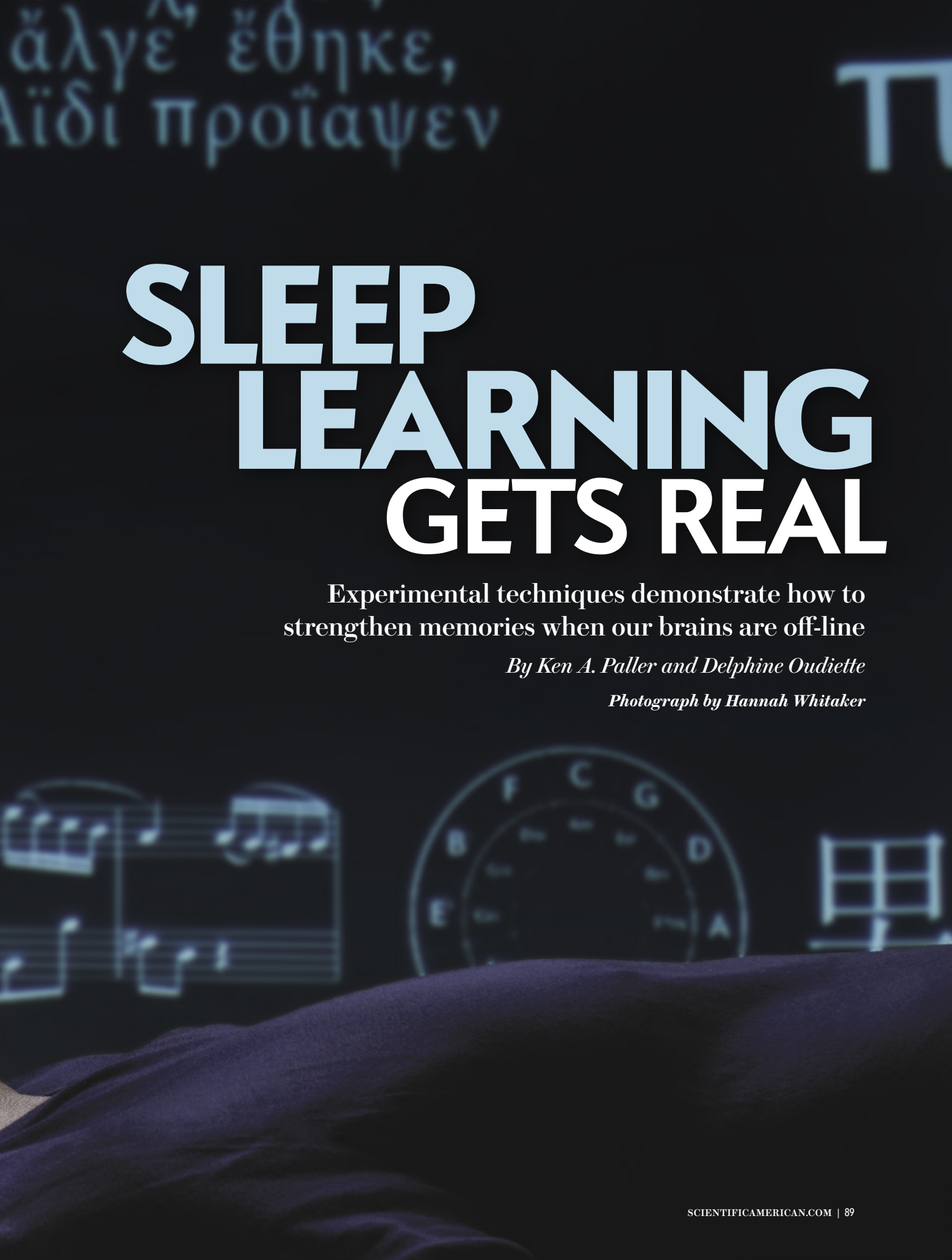
Therapists may need to step in to help certain people. The survey information considered in this article does not delve into nightmares in detail. But some health-care workers who saw relentless suffering later themselves suffered with recurrent nightmares. And some patients who endured the ICU for days or weeks suffered from horrific nightmares during that time, which may in part have been the result of medications and sleep deprivation induced by around-the-clock hospital procedures and interminable monitor noises and alarms. These survivors will need expert help to regain normal sleep. Thankfully, specialized techniques are highly effective.

People who are not traumatized but still a little freaked out about their COVID dreams also have options. New technologies such as targeted memory reactivation are providing individuals with more control over their dream narratives. For example, learning how to practice lucid dreaming—becoming aware that you are dreaming—aided by targeted memory reactivation or other methods could help transform worrisome pandemic dreams into more pleasant, maybe even useful, dreams. Simply observing and reporting pandemic dreams seems to positively impact mental health, as Natália Mota of the Federal University of Rio Grande do Norte in Natal, Brazil, found in her studies.

Short of therapy, we can give ourselves permission to ease up and to enjoy banking those surplus hours of sleep. Dreams can be vexing, but they are also impressionable, malleable and at times inspirational. SA

Tore Nielsen is a professor of psychiatry at the Université de Montréal and director of the Dream and Nightmare Laboratory there.


βλομένην, ἣ μυρί' Ἀχαιοῖς
πολλὰς δ' ἰφθίμους ψυχὰς
$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
Comment allez-vous?



SLEEP LEARNING GETS REAL

Experimental techniques demonstrate how to strengthen memories when our brains are off-line

By Ken A. Paller and Delphine Oudiette

Photograph by Hannah Whitaker

IN ALDOUS HUXLEY'S *BRAVE NEW WORLD,* A BOY MEMORIZES EACH WORD OF A LECTURE in English, a language he does not speak. The learning happens as the boy sleeps within earshot of a radio broadcast of the lecture. On awakening, he is able to recite the entire lecture. Based on this discovery, the totalitarian authorities of Huxley's dystopian world adapt the method to shape the unconscious minds of all their citizens.

Sleep learning turns up throughout literature, pop culture and ancient lore. Take Dexter, the lead character in the animated television series *Dexter's Laboratory*. In one episode, Dexter squanders his time for homework, so instead he invents a contraption for learning to speak French overnight. He wakes up the next day unable to speak anything but French. The idea of sleep learning isn't just a modern invention. It also appears within a centuries-old mind-training practice of Tibetan Buddhists; a message whispered during sleep was intended to help a monk recognize the events in his dreams as illusory.

Everyone knows we learn better when we are well rested. Most people, however, dismiss the notion of sleep learning out of hand. Yet a set of new neuroscientific findings complicates this picture by showing that a critical part of learning occurs during sleep: recently formed memories resurface during the night, and this playback can help reinforce them, allowing at least a few to be remembered for a lifetime.

Some studies have even explored whether sleep might be manipulated to enhance learning. They reveal that sleep's program for making daytime memories stronger can be boosted using sounds and odors. Results in rodents have even demonstrated a primitive form of memory implantation: using electrical stimulation while animals slept, researchers taught them where they should go in their enclosures on awakening. Huxley's imagined version of sleep education, in which entire texts are absorbed verbatim during the night, is still relegated to the pages of his 1932 classic. But experiments now indicate that it is possible to tinker with memories while a person is immersed in the depths of slumber, creating the basis for a new science of sleep learning.

THE PSYCHOPHONE

FOR THESE TECHNIQUES to work, scientists have to explore how information can be absorbed when consciousness is seemingly on a well-deserved break. Around the time that Huxley was writing *Brave New World,* serious explorations into the possibility of meddling with sleep had begun. In 1927 New Yorker Alois B. Saliger invented an "Automatic Time-Controlled Suggestion Machine," which he marketed as the "PsychoPhone," to allow a recorded message to be replayed during the night. The setup seemed to evoke Huxley's imagined technology except that the user, rather than the state, could select the message to be played.

Saliger's invention was followed, in the 1930s and 1940s, by studies documenting ostensible examples of sleep learning. A 1942 paper by Lawrence LeShan, then at the College of William & Mary, detailed an experiment in which the researcher visited a summer camp where many of the boys had the habit of biting their fingernails. In a room where 20 such boys slept, LeShan used a portable phonograph to play a voice repeating the sentence "My fingernails taste terribly bitter." The string of words recurred 300 times each night, beginning 150 minutes after the onset of sleep. The experiment continued for 54 consecutive nights. During the last two weeks of camp, the phonograph broke, so the intrepid LeShan delivered the sentence himself. Eight of the 20 boys stopped biting their nails, whereas none of 20 others who slept without exposure to the recording did so. These early efforts did not use physiological monitoring to verify that the boys were really asleep, though, so the results remain suspect.

The whole field took a severe hit in 1956, when two scientists at RAND Corporation used electroencephalography (EEG) to record brain activity while 96 questions and answers were read to sleeping study participants. (One example: "In what kind of store did Ulysses S. Grant work before the war?" Answer: "A hardware store.") The next day correct answers were recalled only for information presented when sleepers showed signs of awakening. These results led to a shift in the field that persisted for 50 years, as researchers began to lose faith in sleep learning as a viable phenomenon: participants in these experiments appeared to learn only if they were not really sleeping when information was presented to them.

Most scientists during this time tended to avoid the topic of sleep learning, although a few researchers did

plug away at asking whether sleep assists in remembering new information. One typical study protocol probed whether overnight sleep deprivation affected recall the day after learning something new. Another asked whether remembering was better after a nap than after an equal period of time spent awake.

Various confounding factors can interfere with such studies. For example, the stress of sleep deprivation can harm cognitive functions, which then decreases memory recall. Eventually cognitive neuroscientists began to tackle these challenges by bringing together evidence from multiple research methods. A substantive foundation of evidence gradually accrued to confirm that sleep is a means of reviving memories acquired during the day, reopening the relation between sleep and memory as a legitimate area of scientific study.

Many researchers who took up the challenge focused on rapid eye movement (REM) sleep, the period when dreams are the most frequent and vivid. The guiding assumption held that the brain's nighttime processing of memories would be tied to dreaming, but clear-cut data did not materialize. In 1983 two noted scientists—Graeme Mitchison and Francis Crick, neither a psychologist—went so far as to speculate that REM sleep was for forgetting. In a similar vein, Giulio Tononi and Chiara Cirelli, both at the University of Wisconsin–Madison, proposed that sleep could be the time for weakening connections among brain cells, making it easier for new information to be acquired the following day.

Instead of REM, some investigators focused their attention on slow-wave sleep (SWS), a period of deep slumber without rapid eye movements. In 2007 Björn Rasch, then at the University of Lübeck in Germany, and his colleagues prepared people for a sleep experiment by requiring them to learn the locations of a set of objects while simultaneously smelling the odor of a rose. Later, in their beds in the laboratory, sleeping study participants again encountered the same odor as electrical recordings confirmed one sleep stage or another. The odor activated the hippocampus, a brain area critical for learning to navigate one's surroundings and for storing the new knowledge gained. On awakening, participants recalled locations more accurately—but only following cueing from odors that emanated during the course of slow-wave (not REM) sleep.

TARGETED MEMORY REACTIVATION

IN 2009 OUR LAB EXTENDED this methodology by using sounds instead of odors. We found that sounds played during SWS could improve recall for individual objects of our choosing (instead of the recall of an entire collection of objects, as was the case in the odor study). In our procedure—termed targeted memory reactivation, or TMR—we first taught people the locations of 50 objects. They might learn to place a cat at one designated spot on a computer screen and a teakettle at another. At the same time, they would hear a corresponding sound (a meow for the cat, a whistle for the kettle, and so on).

After this learning phase, participants took a nap in a comfortable place in our lab. We monitored EEG recordings from electrodes placed on the head to verify that each individual was soundly asleep. These recordings provided intriguing data on the synchronized activity of networks of neurons in the brain's outer layer, the cerebral cortex, that are relevant for memory reactivation [*see box on next page*]. When we detected slow-wave sleep, we played the meow, whistle and other sounds associated with a subset of the objects from the learning phase. Sounds were presented softly, not much louder than background noise, so the sleeper did not awaken.

On awakening, people remembered locations cued during sleep better than locations that had not been cued during sleep. Whether sounds or odors served as cues in these experiments, they apparently triggered the reactivation of spatial memories and so reduced forgetting.

Future programs for sleep learning might help in preserving memories, speeding acquisition of new knowledge, or even changing bad habits such as smoking.

At first, the auditory procedures we used were highly controversial. The received wisdom among sleep researchers held that sensory circuits in the cortex are largely switched off during sleep, except for the sense of smell. We were not swayed by this orthodox view. Instead we followed our hunch that the repeated playing of soft sounds might influence the sleeping brain and produce changes in recently stored memories.

Indeed, the same memory benefits were also found in many subsequent studies. A technique called functional magnetic resonance imaging highlighted which brain areas take part in TMR, and EEG results brought out the importance of specific brain oscillations. Two papers published in 2018—one by Scott Cairney of the University of York in England and his colleagues, the other by James Antony of Princeton University and his colleagues—linked an oscillation, the sleep spindle, with the memory benefits of TMR.

Besides boosting spatial memory, these procedures have also helped improve recall in other settings. TMR can assist in mastery of playing a keyboard melody and learning new vocabulary or grammatical rules. The technique can also help with simpler types of learning, such as adjustments in one's body image. In conditioning experiments, TMR alters prior learning of an automatic reaction to a stimulus caused by an earlier pairing

The Maestros of Slumber

A complex symphony of neural activity governs the connection between sleep and memory

Brain rhythms provide clues to how sleep helps to store memories for later retrieval. One type of neural signal, called a slow wave, cycling from 0.5 to four times a second, orchestrates the activity of neurons in the cerebral cortex. Each slow oscillation consists of a "down" phase, when neurons are silent, and an "up" phase, when they resume activity. This timing pattern helps to reinforce recently formed memories by ensuring that multiple cortical regions remain in an up state at the same time.

The up phase can coincide with sleep spindles, brief increases of a rhythm of 12 to 15 cycles per second. Spindles originate in the thalamus, which serves as a crossroads for information that is transmitted to virtually all parts of the cerebral cortex. Spindles have a rhythm of their own, recurring at approximately five-second intervals. They coordinate the activity of sharp-wave ripples in the hippocampus. Ripples, for their part, are concurrent with the replay of memories. Slow waves, all the while, assume the role of orchestra conductor: their measured oscillations in the cortex coordinate the pacing for sleep spindles and sharp-wave ripples.

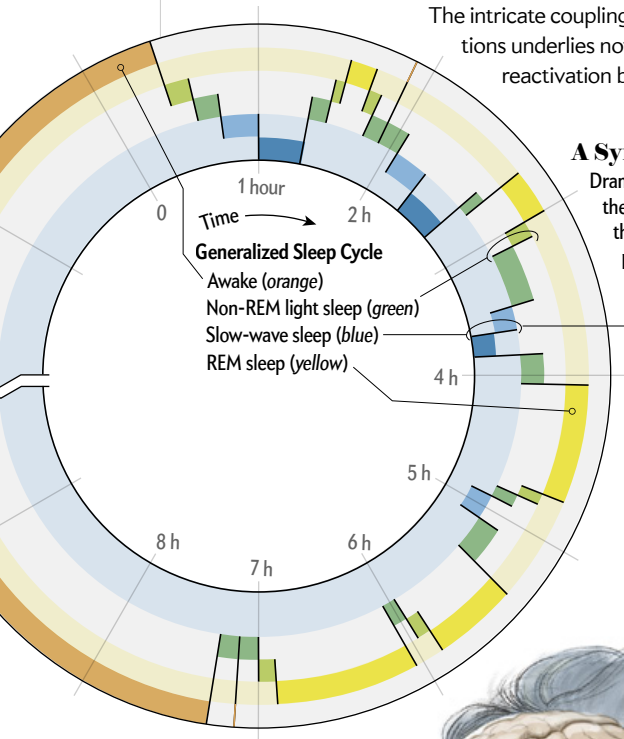
The intricate coupling of these oscillations underlies not only memory reactivation but also the altering of connections among neurons to strengthen memory storage. A dialogue between the hippocampus and the cortex involving all these brain rhythms triggers a set of complex network interactions. Through this process, known as consolidation, new information can become integrated with existing memories. The intertwining of memories, moreover, enables the gist of recent experiences to be extracted to make sense of a complex world.

Memory difficulties can arise when this neural dialogue becomes impaired. Individuals with major damage centered in the hippocampus or parts of the thalamus may develop a profound amnesia. Without the expected interactions with these brain regions during both sleep and waking, the cortex cannot store mental records of facts and events known as declarative memories. In addition, a milder form of memory disorder may result when memory processing during sleep is seriously disrupted.

Deciphering the physiological orchestration of the sleeping brain is prompting various new strategies for enhancing the brain's natural rhythms—including stimulation with slow electrical oscillations, sounds or gentle motion. These methods echo humans' natural inclinations to take advantage of a lullaby's rhythm or the rocking of a cradle to lull a baby to sleep. *—K.A.P. and D.O.*

A Symphony in Two Movements

Dramatic differences characterize two key sleep phases. The slow waves of deep sleep dominate the early part of the night. During slow-wave sleep, some memories are spontaneously reactivated. Interventions that promote this process can ensure that memories are retained. Rapid eye movement (REM) sleep prevails in the latter part of a night's slumber, but how it interacts with memory remains controversial.



Harmonizing Brain Waves

Brain oscillations during sleep appear to play a role in strengthening new memories. A key event is the "up" phase of a slow oscillation that coordinates the activity of other brain rhythms. The ascending part of a slow oscillation in the cortex synchronizes with sleep spindles in the thalamus. The spindles coordinate the activity of sharp-wave ripples in the hippocampus. Ripples tend to coincide with a spindle trough.

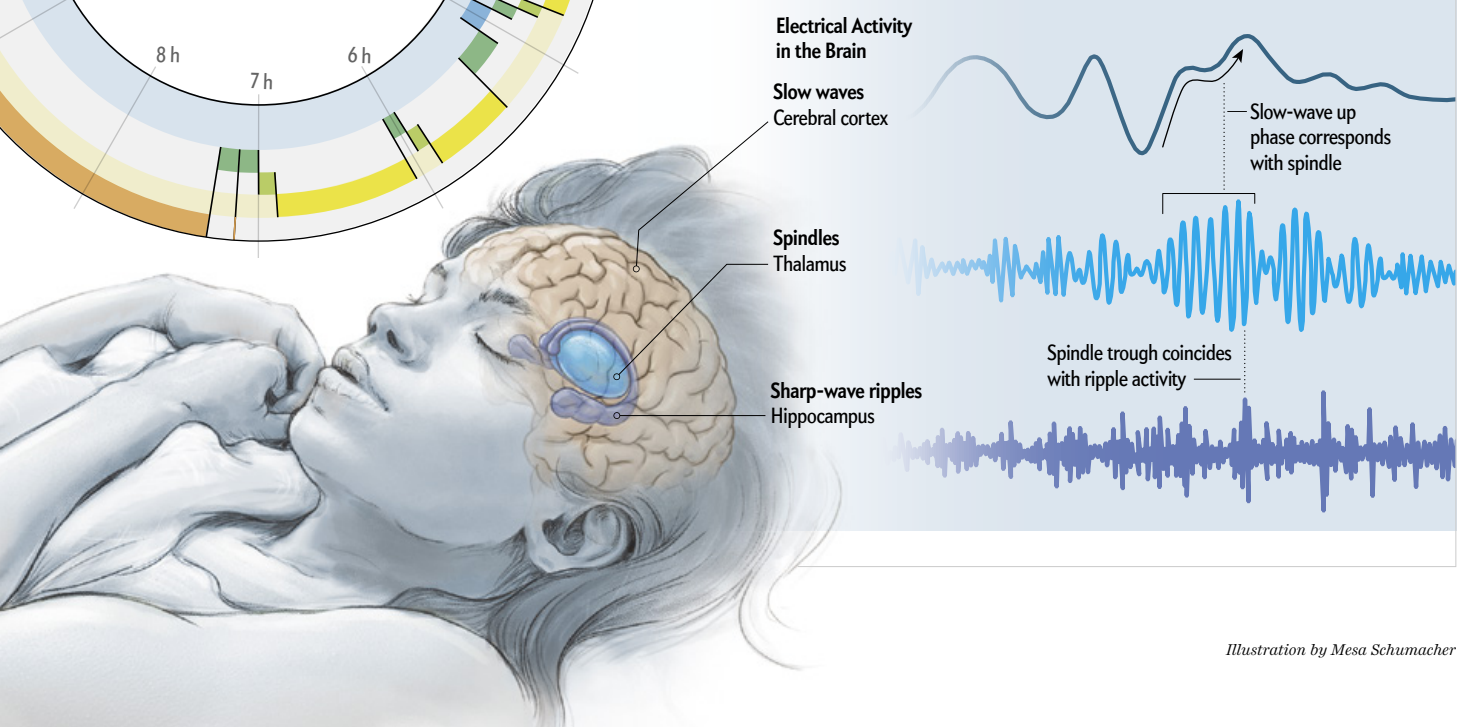


Illustration by Mesa Schumacher

of that stimulus with an electric shock. Ongoing studies are examining still other types of recall, such as associating names with faces when first meeting new people.

As the technology evolves, TMR should be tested to see whether it could help to treat various disorders, reduce addictions or speed recovery from illness. Our lab, together with Northwestern University neurologist Marc Slutzky, is currently testing a novel rehabilitation procedure for recovering arm-movement abilities after stroke. Cue sounds are incorporated as part of the therapy and are replayed during sleep to try to accelerate relearning of lost movements. The prospects appear promising because TMR can alter similar forms of motor learning in healthy individuals.

WHAT ABOUT LEARNING FRENCH?

THE DEMONSTRATED ABILITY to reinforce memories raises the question of whether new information can be loaded into a person's brain after falling asleep, a technique that calls forth the ethical specter of mind control invoked by *Brave New World*. Is it going too far, though, to imagine that memories can be created surreptitiously?

Although the orthodox response to such conjectures has for many years been an unqualified no, studies by Anat Arzi, then at the Weizmann Institute of Science in Rehovot, Israel, and her colleagues demonstrated the creation of relatively simple memories using odors. In one experiment, the researchers succeeded in diminishing the desire for tobacco in smokers who were keen to quit. When asleep, study participants were exposed to two odors, cigarette smoke and rotten fish. During the next week, those who had smelled the mix of both odors lit up 30 percent less, having apparently been conditioned to associate smoking with the aversive fish odor.

Acquiring a more complex memory is not as easy, but evidence from the past decade holds tantalizing promise. Karim Benchenane of the French National Center for Scientific Research (CNRS) and his colleagues have shown how to literally change the mind—of a mouse. When they began their work, Benchenane and his team knew that when a mouse explores a new environment, neurons called place cells fire as the animal traverses specific parts of an enclosure. These same neurons discharge again during sleep as the memory is apparently replayed.

The researchers stimulated the reward system of the mouse brain (the medial forebrain bundle) precisely when place cells became spontaneously active while the animal was asleep. Amazingly, mice subsequently spent more time at the locations that corresponded to the stimulated place cells, heading there directly after they woke up. More experiments still need to disentangle whether fully formed false memories were implanted in the mice during sleep or whether they were automatically guided to those spots by a process of conditioning, without more knowledge about why they were drawn to those locations.

In 2019 Swiss researchers reported that sleepers could acquire new verbal knowledge, but this was evident only through subtle nonconscious means. More recently, we showed episodic learning during sleep with full recollection of the learning. In a multi-institutional collaboration by researchers in France, Germany, the Netherlands and the U.S., we used a variant of the TMR method to encourage lucid dreaming—a state in which people realize they are dreaming while remaining in the midst of the dream. We then showed that people could understand softly spoken questions from within these dreams and produce correct answers by signaling with their eyes, their respiration or subtle twitches in their facial muscles. Sometimes people in these experiments woke up able to recollect parts of their dreamtime Q and A. These rare occurrences convincingly document full-blown learning experienced entirely during sleep.

The boundaries of sleep hacking may continue to expand, but this research has established that a normal component of learning continues nocturnally off-line. Sleep is needed not just to help people stay alert and rejuvenated but also to reinforce memories initially acquired while they were awake. We still need to learn much more about off-line memory processing. Further work must ascertain how sleep helps learning and which brain mechanisms are engaged to preserve the most valuable memories. It is also essential to find out more about the perils of poor or inadequate sleep that might be affected by various forms of life stress, certain diseases or the experience of growing older.

A study led by Carmen Westerberg, then at Northwestern, points in the desired direction. Westerberg tested patients with the memory dysfunction that often precedes Alzheimer's disease—amnestic mild cognitive impairment. The results documented a link between poor sleep and reduced ability to remember information after an intervening overnight delay.

All of this knowledge might help in creating programs of sleep learning to preserve memories, to speed the acquisition of new knowledge, or even to change bad habits such as smoking. Looking still further ahead, scientists might also explore whether we can gain control over our dreams, which could lead to the prospect of nightmare therapies, sleep-based problem-solving and perhaps even recreational dream travel. In a culture that already offers wrist-based sleep trackers and mail-order gene tests, we can begin to contemplate new ways to convert daily downtime into a productive endeavor—for some, a chilling prospect, and for others, another welcome opportunity for self-improvement. SA

Ken A. Paller is a professor of psychology, holds the James Padilla Chair in Arts & Sciences, and directs the cognitive neuroscience program at Northwestern University. His research on targeted memory reactivation was funded by the U.S. National Science Foundation.

Delphine Oudiette is a tenured researcher for the French National Institute for Health and Medical Research (INSERM) at the Brain & Spine Institute and at the sleep disorder department located at Pitié-Salpêtrière Hospital, both in Paris.

Answering Queries in Real Time while Dreaming

Researchers demonstrate that during REM sleep, people can hear—and respond to—simple questions such as "What is eight minus six?"

By Diana Kwon

Dreams are full of possibilities; by drifting into the world beyond our waking realities, we can visit magical lands, travel through time and interact with long-lost family and friends. The notion of communicating in real time with someone outside of our dreamscapes, however, sounds like science fiction. A recent study demonstrates that, to some extent, this seeming fantasy can be made real.

Scientists already knew that one-way contact is attainable. Previous studies have demonstrated that people can process external cues, such as sounds and smells, while asleep. There is also evidence that people are able to send messages in the other direction: lucid dreamers—those who can become aware they are in a dream—can be trained to signal, using eye movements, that they are in the midst of a dream.

Two-way communication, however, is more complex. It requires a person who is asleep to actually understand what they hear from the outside and think about it logically enough to generate an answer, explains Ken Paller, a cognitive neuroscientist at Northwestern University. "We believed that it was going to be possible—but until we actually demonstrated it, we weren't sure."

For this study, Paller and his colleagues recruited volunteers who said they remembered at least one dream per week and provided them with guidance on how to lucid dream. They were also trained to respond to simple math problems by moving their eyes back and forth—for example, the correct response to "eight minus six" would be to move your eyes to the left and right twice. While the participants slept, electrodes attached to their faces picked up their eye movements, and electroencephalography (EEG)—a method of monitoring brain activity—kept track of what stage of sleep they were in.

As Paller's team was conducting these experiments, the researchers discovered three groups in Germany, France and the Netherlands who were trying to accomplish the same thing. Instead of competing, the groups decided to collaborate. They carried out similar experiments, though with slightly different methods of answering questions and receiving responses. The German group, for example, transmitted its math problems using Morse code, and the French group asked its participant—a person with narcolepsy who had expert lucid-dreaming abilities—to answer yes-or-no questions with facial muscle contractions rather than eye movements.

Across the four studies, there were a total of 36 participants and 158 trials during which the researchers could verify lucid dreaming and attempted to establish contact. Answers were considered correct if three of four raters were in agreement on whether the responses, sometimes very subtle movements, were accurate. Correct responses were given in 18 percent of trials; another 18 percent were classified as ambiguous because raters could not come to a consensus about whether participants gave a correct response or whether they had responded at all. Incorrect responses were given in 3 percent of the trials. Overall there was no response in 60 percent.

One of the co-authors, Karen Konkoly, a graduate student in Paller's lab, speculates that participants failed to respond in 60 percent of the trials because they simply did not perceive the incoming communication. In those cases, they rarely reported any incorporation of the questions into their dreams after waking up. But

Simon Kadula/Alamy

she adds that it is also possible that dreamers perceived the inputs but paid little attention and forgot before awakening. The proportion of people who respond could potentially be improved with more training or by presenting questions when individuals are in specific sleeping brain states, Konkoly says.

After establishing successful two-way communication, participants were woken up and asked to recount their dreams. In most cases, they could remember receiving the experimenters' questions while asleep; in some cases, the questions appeared to be coming from outside the dream, whereas other times they were integrated into the dream. (One participant reported that the lights in their dream started flickering, which they were able to recognize as the Morse-coded math problem.)

There were instances, however, when people either did not recall the interactions or had a distorted account. For example, there were trials in which individuals answered a math problem correctly while asleep but did not remember the question correctly after waking up. These findings were published in February 2021 in *Current Biology*.

The findings "challenge our ideas about what sleep is," says Benjamin Baird, a researcher who studies dreams at the University of Wisconsin–Madison and was not involved in this study. Sleep has classically been defined as unresponsiveness to external environmental stimuli—and that feature is still typically part of the definition today, Baird explains. "This work pushes us to think carefully—rethink, maybe—about some of those fundamental definitions about the nature of sleep itself and what's possible in sleep."

This kind of two-way communication with dreamers could be used as a tool to better study dreams, according to Paller. In particular, he says, the observation that the responses some people gave during dreams did not match their reports after waking provides evidence that such real-time techniques will help researchers get more accurate accounts of dreams—and address whether dreams play a useful role in processes such as memory. Paller and his colleagues also suggest this technique could be used by people to enhance problem-solving and creativity, by providing a new way to process content in their dreams.

"I really liked this study," says Christine Blume, a sleep scientist at the Center for Chronobiology in Basel, Switzerland, who was not involved in this work. "The extent to which information can be processed and responded to surprised me." But she adds that it is important to keep in mind that the findings relate specifically to lucid dreaming, which is a special type of dreaming that not many people are able to experience.

Blume notes that even with lucid dreamers, in most trials, the researchers were not able to establish communication. Therefore, how applicable this technique would be to learning or creativity remains an open question, she says.

Paller and his colleagues are now exploring what other types of questions can be asked during sleep, as well as other ways of receiving messages from sleepers, such as sniffing. "We are hopefully going to get better at doing this kind of experiment," he observes. "Then [we can] ask new questions about what's happening during dreams." SA

Diana Kwon is a freelance journalist who covers health and the life sciences. She is based in Berlin.

OPINION

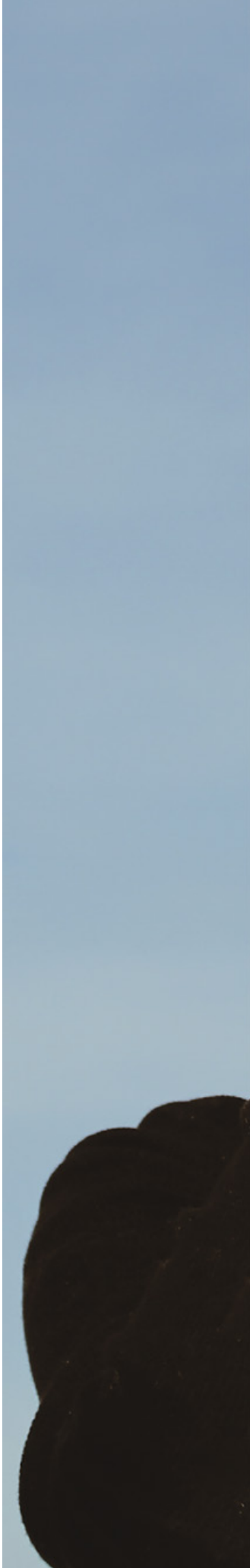
CAN SCIENCE ILLUMINATE OUR INNER DARK MATTER?

Neither introspection nor brain scans can reveal our deepest thoughts

By John Horgan

WHAT'S GOING ON IN YOUR HEAD RIGHT NOW? HOW ABOUT … now? Or … now? Answering this question is harder than you might think. As soon as you pay attention to your thoughts, you alter them, as surely as you alter an electron's course by looking at it. You can't describe your thoughts the way you describe, say, the room in which you are reading, which remains stolidly unaffected by your scrutiny.

William James draws attention to this paradox in "The Stream of Thought," a section of *The Principles of Psychology*. Trying to examine your thoughts through "introspective analysis," he writes, is like studying snowflakes by catching them in your "warm hand," "seizing a spinning top to catch its motion" or "trying to turn up the gas quickly enough to see how the darkness looks."



Maria Magliocco/Eye Em/Getty Images

I've been brooding over these sorts of puzzles a lot lately, even more than usual. In 2020 I released a book entitled *Pay Attention: Sex, Death, and Science,* a stream-of-thought account of a day in my life. Or rather, in the life of my fictional alter ego, Eamon Toole, a "divorced science writer and professor struggling to remain rational while buffeted by fears and desires." As soon as I send a book out into the world, I compulsively think of things I should have put in it.

Also, my ongoing attempt to learn quantum mechanics has mystified my world, inner and outer. So, I'd like to offer a few thoughts (second thoughts? afterthoughts?) about thoughts, the most inescapable and maddeningly elusive features of our existence.

META-THOUGHTS AND THOUGHTLESS THOUGHTS

A NOTE ON TERMS. James coined the phrases "stream of thought" and "stream of consciousness" and sometimes used them interchangeably, but I distinguish thoughts from consciousness in the following way: Thoughts are the *contents* of consciousness, including fears, fantasies, recollections, realizations, deliberations, decisions and all the other flora of subjective experience. If consciousness is the medium, thoughts are the message.

I also like the easy self-referentiality of "thoughts about thoughts," which captures a deep truth about us. We are what Douglas Hofstadter calls self-generating "strange loops," akin to M. C. Escher's famous drawing of two hands drawing each other. (Who draws the drawer?) Thoughts spring from thoughts and—in ways still beyond our ken—from our brains, which contain roughly 100 billion neurons linked by one quadrillion synapses, each of which processes an average of 10 electrochemical signals, or action potentials, every second.

THOUGHTS ARE "CONTINUOUS," THEY "FLOW," THEY KEEP COMING EVEN WHEN WE PAY NO ATTENTION TO THEM, AND THEY KEEP CHANGING; NO THOUGHT IS PRECISELY LIKE ANOTHER.

If you equate action potentials with the operations of a computer, as many neuroscientists do, then the brain carries out 10 quadrillion operations in a typical second. That approaches the speed of the world's fastest supercomputers, and the brain may perform exponentially more calculations via processes other than action potentials. The result of all this activity is that brains churn out thoughts as ceaselessly as hearts pump blood.

As James puts it, thoughts are "continuous," they "flow," they keep coming even when we pay no attention to them, and they keep changing; no thought is precisely like another. James thus doubts whether psychologists can reduce the human mind to a mental equivalent of atoms as physicists have done with matter.

With some effort, I can direct my thoughts, focus them, but they often seem to have a will of their own. They swerve this way or that for reasons obscure to me, a tendency that Buddhists disparage as "monkey mind." When we do notice a thought and reflect on it—perhaps to convey it to ourselves or to others—we instantly transform it, turning it into a different, higher-order thought. Call it a meta-thought, a thought about a thought.

Meta-thoughts—the thoughts I express to myself and to others through writing and speech—are my bread and butter. I make my living off them. But they constitute an infinitesimal fraction of my thoughts. The vast majority are unformed, incoherent, inexpressible, and they come and go without my dwelling on them. You might call them thoughtless thoughts. Thoughtless thoughts are what course through your head when no one is watching you, not even you.

CONSCIOUSNESS WITHOUT THOUGHT?

I ONCE TRIED TO TEACH meditation to a class of stressed-out freshmen. I told them to close their eyes, still their minds and stop thinking. After 10 minutes of silence, I asked how many had succeeded in thinking of nothing. To my surprise, about half raised their hands. I didn't believe them. Even freshmen always have thoughts, whether or not they notice them.

Is thoughtless consciousness possible? Yes, according to religious scholar Robert Forman, a veteran meditator. He claims that he and others have achieved "pure consciousness," a mystical state devoid of any specific thoughts. You are conscious but not conscious *of* anything. Consciousness without content strikes me as a contradiction, an oxymoron, like a book without words or a film without images. And how would you *know* you're in a state of pure consciousness? How would you remember it? Even Forman admits that states of pure consciousness, if they exist, are rare.

Meditation is touted as a route to knowledge of your deepest self, your innermost thoughts. I've had delightful experiences meditating, especially on a silent retreat in 2018. But meditation and other contemplative techniques are designed to control and suppress thoughts rather than to understand them. Meditation is self-brainwashing aimed at taming your monkey mind. I don't want to tame my monkey mind; I want to study it, to comprehend its antics.

Although we may not notice them and may even deny their existence, thoughtless thoughts are always there, underpinned by our brains' incessant chatter. Without thoughtless thoughts, we would lack meta-thoughts. Thoughtless thoughts are the dark matter of our minds, giving shape via hidden mechanisms to what is observable, visible, luminous in our inner cosmos.

THE LIMITS OF STREAM-OF-THOUGHT FICTION

CAN WE STUDY THOUGHTLESS thoughts, the mind's dark matter, given that simple introspection doesn't work? Some neuroscientists predict that external brain-scanning devices, such as MRIs, or arrays of implanted electrodes will soon allow us to read minds. But this feat would require cracking the neural code, the set of rules or algorithms that turn neural activity into mental activity—that is, thoughts. The neural code is the enigma at the core of the mind-body problem. The more we investigate it, the more intractable it seems.

I try to describe my thoughts in my book *Pay Attention* (original title: *What Is It Like to Be a Science Writer?*). The book is based on journals in which I wrote down what I did, saw, said, heard and thought over the course of a typical day, as I commuted to the university where I work, talked to a freshman humanities class (about James's "Stream of Thought"), jawed with colleagues over lunch (about Thomas Kuhn's views of "truth") and spent the evening with "Emily," my girlfriend.

In the first draft of my book, to make my thoughts seem more raw and real, I expressed them as sentence fragments running into each other, with little punctuation. An editor who read this draft described it as "sludge." Even I found that draft hard to read. So I cheated. I rewrote the book with more or less coherent sentences with conventional grammar and punctuation. I also added contextual information for readers, information that I wouldn't have actually thought about because I just knew it implicitly.

To justify these moves in the direction of readability, I could point out that the book's narrator is an extremely self-conscious science writer trying to make his private thoughts public. He is in a sense performing his thoughts, first for himself and then for readers. But that means my book consists of meta-thoughts. It isn't an accurate depiction of my thoughtless thoughts, which remain veiled from me.

Nobody depicts thoughts in all their raw weirdness as vividly as James Joyce. In *Ulysses,* Joyce plops us inside the heads of Stephen Dedalus, a teacher and aspiring writer and an avatar for Joyce as a young man; Leopold Bloom, a nerdy, genial ad salesman; Bloom's voluptuous wife, Molly; and other characters living in Dublin in the early 20th century. We see, feel, remember what they see, feel, remember.

But Joyce's notoriously difficult masterpiece isn't entirely stream of thought. If it were, it would be far more difficult. To help orient us, give us a little context, Joyce occasionally shifts his point of view from inside characters' heads to outside, that is, from a first-person to a third-person perspective.

Joyce's final opus, *Finnegans Wake,* which I "read" in college, makes no concessions to readability. Even Joyce's admirers complained about its opacity, but James defended his gobbledygookian work. "One great part of every human existence," he told a friend, "is passed in a state which cannot be rendered sensible by the use of wideawake language, cutanddry grammar and goahead plot." But even *Finnegans Wake,* an unrivaled imagining of mental dark matter, consists entirely of Joyce's hyperconscious, insanely erudite meta-thoughts. And what about all the thoughts that cannot be captured by words?

WILL SCIENCE EVER DISCOVER A FINAL THEORY OF THE MIND? ONE THAT SOLVES THE MIND-BODY PROBLEM AND MAKES US FULLY TRANSPARENT TO OURSELVES?

HIDDEN-VARIABLE THEORIES OF THE MIND

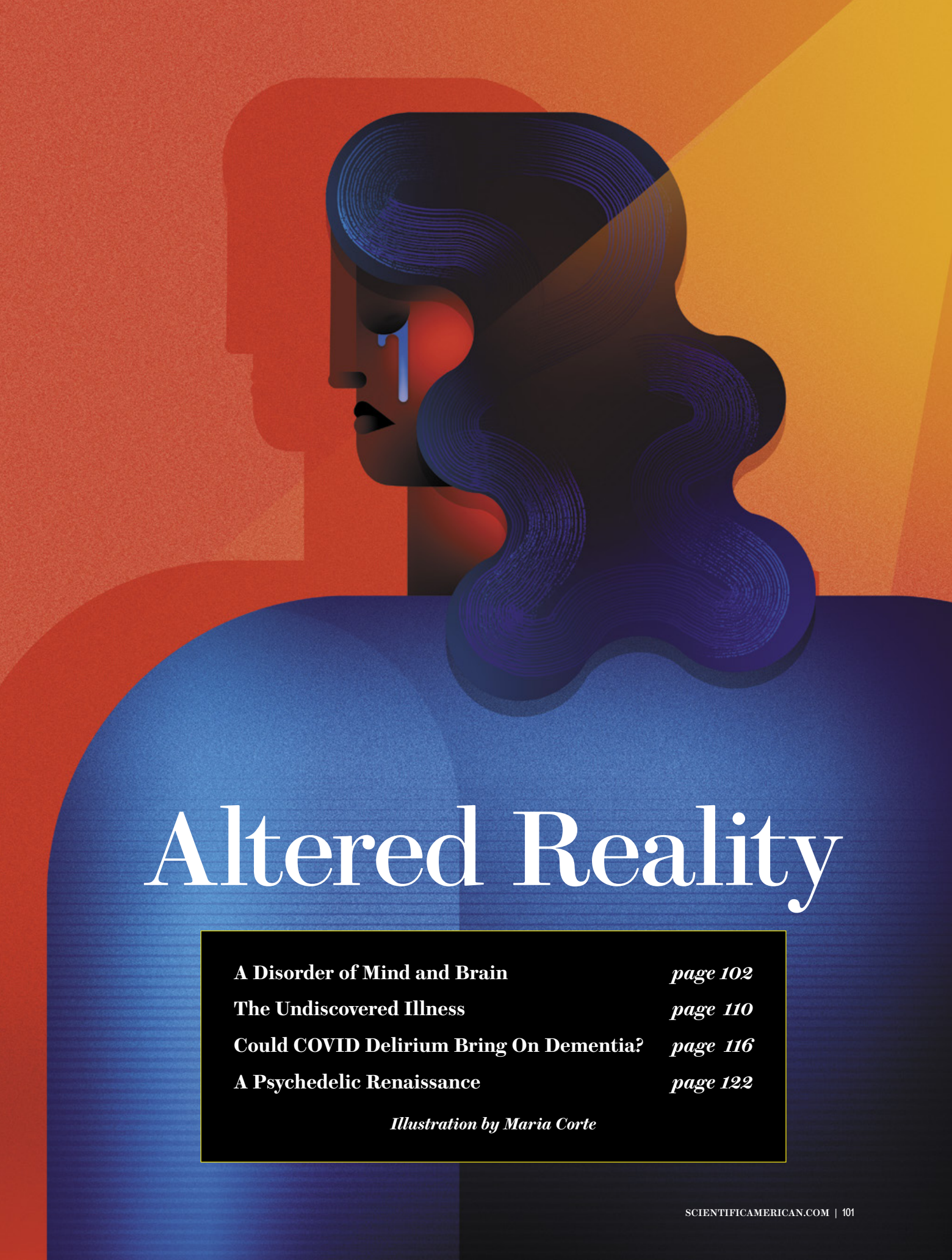
A FINAL POINT: I see analogies between efforts to understand thoughts and the quantum realm. I alluded to one correlation above: observing particles alters them, as does observing thoughts. Here's another: some physicists, dissatisfied with probabilistic quantum accounts of electrons and photons, seek to explain their behavior in terms of "hidden variables" that follow deterministic rules.

Mind scientists, similarly, have proposed hidden-variable paradigms of the mind. Psychoanalysis holds that our conscious minds are yanked this way and that by deep-rooted lusts and aversions. Evolutionary psychology traces our emotions and actions to instincts embedded in our ancestors by natural selection. Cognitive science postulates that our thoughts stem from computations carried out by our neural machinery and are as far removed from our conscious thoughts as the machine code of your smartphone is from the icons on its screen.

Although each of these paradigms has appealing features, each finally falls short, as do all theories of the mind. Will science ever discover a final theory of the mind? One that solves the mind-body problem and makes us fully transparent to ourselves? That reveals the hidden variables underpinning and linking our meta-thoughts and thoughtless thoughts?

I doubt it. Physicists can't grok the behavior of a single electron that is identical to every other electron. So what hope do we have of capturing the thought passing through your head right … now, a thought unlike any that you, let alone anyone else, have ever had or will have? And if we can't grasp a single thought, which melts the instant we grasp it, how can we possibly understand ourselves? Think about *that.* SA

John Horgan directs the Center for Science Writings at the Stevens Institute of Technology. His books include *The End of Science, The End of War* and *Mind-Body Problems,* available for free at mindbodyproblems.com. For many years he wrote the popular blog Cross Check for *Scientific American.*



Altered Reality

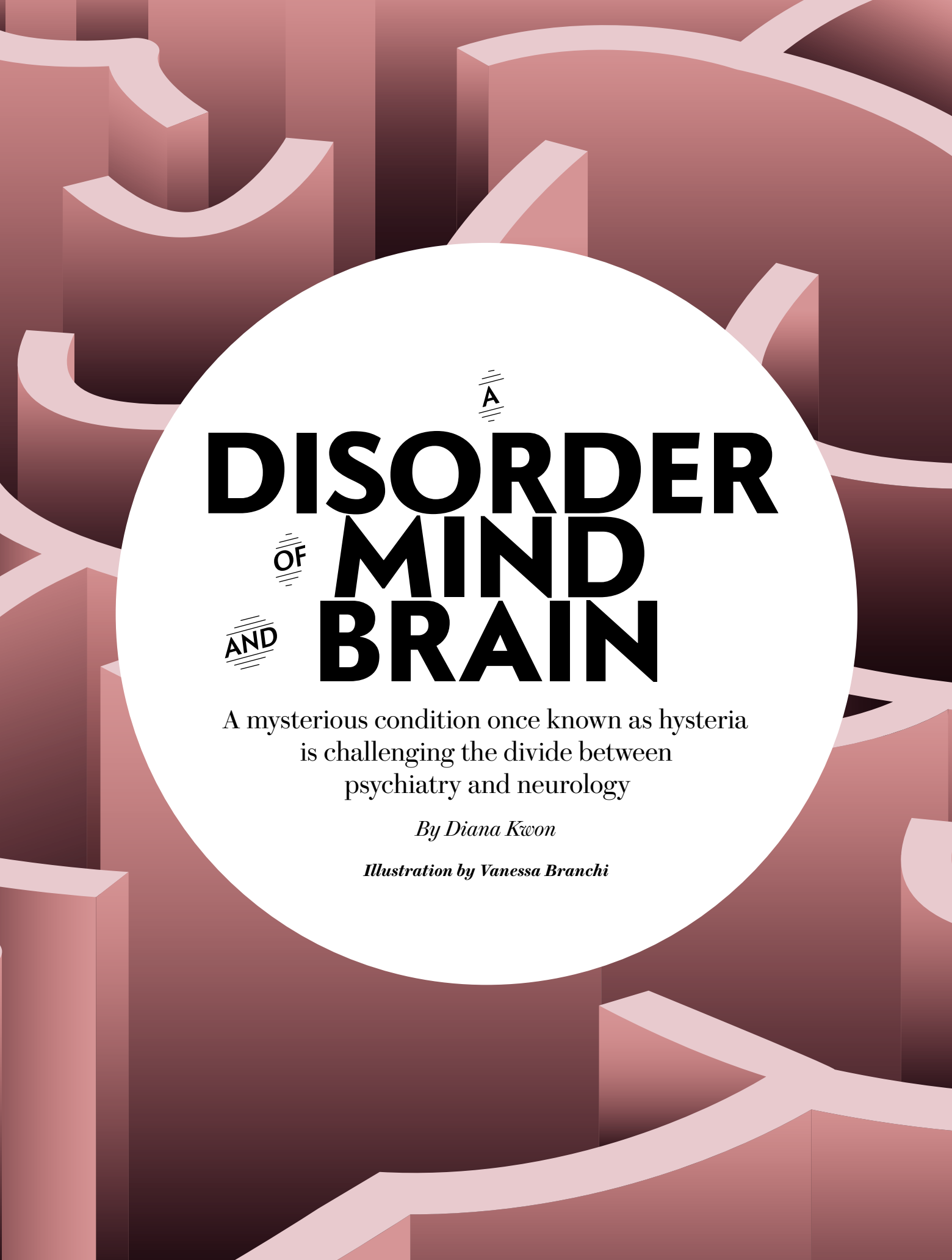
A Disorder of Mind and Brain *page 102*

The Undiscovered Illness *page 110*

Could COVID Delirium Bring On Dementia? *page 116*

A Psychedelic Renaissance *page 122*

Illustration by Maria Corte

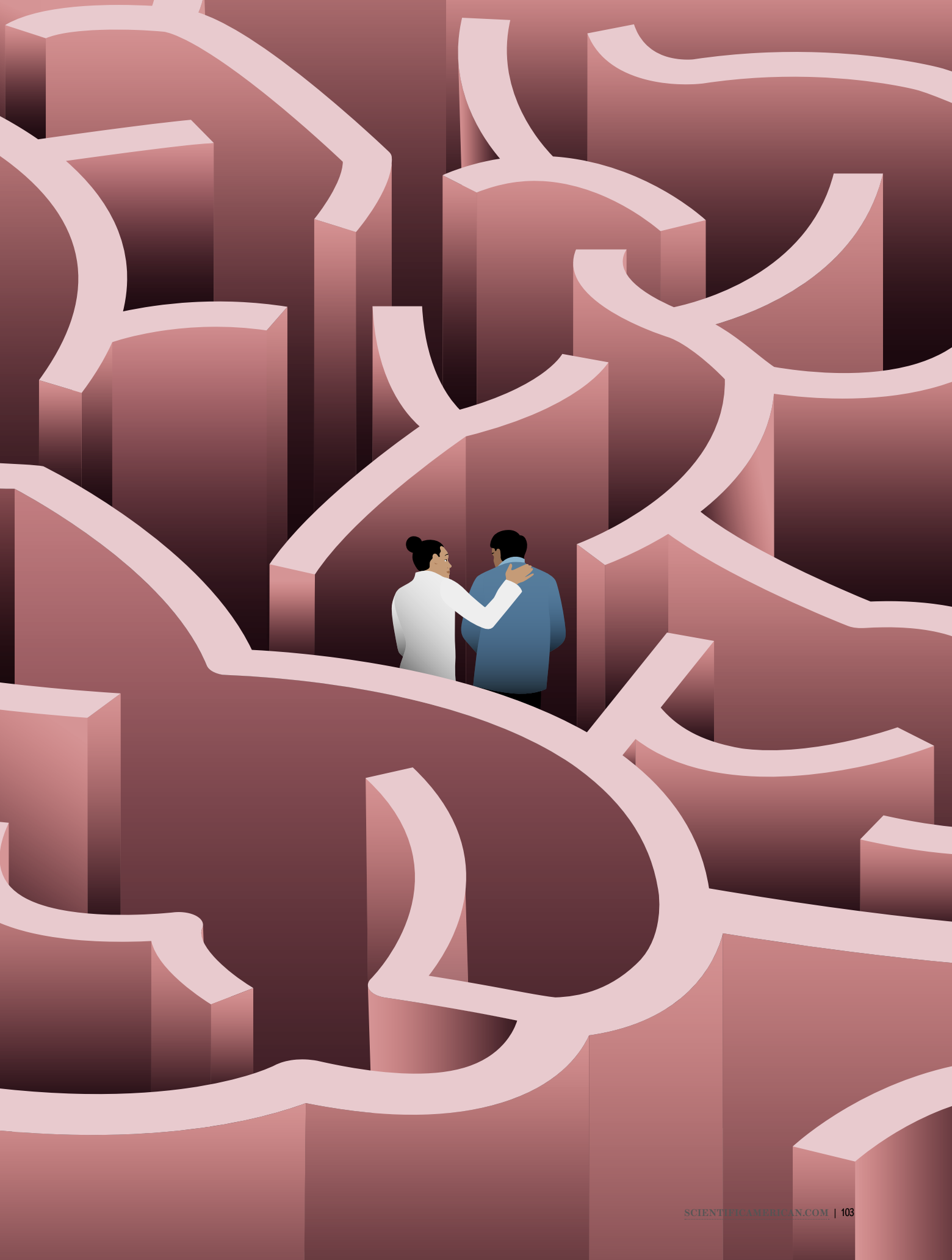


A DISORDER OF MIND AND BRAIN

A mysterious condition once known as hysteria is challenging the divide between psychiatry and neurology

By Diana Kwon

Illustration by Vanessa Branchi



It all began with a cough.

Four years ago Tracey McNiven, a Scottish woman in her mid-30s, caught a bad chest infection that left her with a persistent cough that refused to subside, even after medication. A few months later strange symptoms started to appear. McNiven noticed numbness spreading through her legs and began to feel that their movement was out of her control. When she walked, she felt like a marionette, with someone else pulling the strings. Over the course of two weeks the odd loss of sensation progressively worsened. Then, one evening at home, McNiven's legs collapsed beneath her. "I was lying there, and I felt like I couldn't breathe," she recalls. "I couldn't feel below my waist." McNiven's mother rushed her to the hospital where she remained for more than half a year.

During her first few weeks in the hospital, McNiven endured a barrage of tests as doctors tried to uncover the cause of her symptoms. It could be a progressive neurodegenerative condition such as motor neuron disease, they thought. Or maybe it was multiple sclerosis, a disease in which the body's own immune cells attack the nervous system. Bafflingly, however, the brain scans, blood tests, spinal taps and everything else came back normal.

McNiven's predicament is not uncommon. According to one of the most comprehensive assessments of neurology clinics to date, roughly a third of patients have neurological symptoms that are deemed to be either partially or entirely unexplained. These may include tremor, seizures, blindness, deafness, pain, paralysis and coma and can parallel those of almost any neurological disease. In some patients, such complications can persist for years or even decades; some people require wheelchairs or cannot get out of bed. Although women are more often diagnosed than men, such seemingly inexplicable illness can be found in anyone and across the life span.

Generations of scientists have tried to understand these bizarre conditions, which have historically been given diverse names, such as hysteria, conversion disorder or psychosomatic illness. These labels have, however, long imposed particular explanations for what many researchers now regard as a complex illness at the interface of psychiatry and neurology. Some are still in use today, but the newest name for these conditions, functional neurological disorder (FND), is deliberately neutral, simply denoting a problem in the functioning of the nervous system.

Patients with FND have long struggled to obtain adequate care. They have been accused of feigning or imagining symptoms, painfully but often fruitlessly probed for childhood trauma and dismissed by doctors who did not know how to treat someone who, based on all the usual tests, appeared to be healthy. "For many, many years physicians have underestimated the prevalence of these disorders and the human toll it takes," says Kathrin LaFaver, a neurologist who specializes in movement disorders at Saratoga Hospital in New York State. "These people have really fallen [in the gap] between the fields of neurology and psychiatry."

Over the past decade or so, however, using techniques such as functional magnetic resonance imaging (fMRI), researchers have begun to understand what happens in the brains of patients with this enigmatic illness. And by applying new models of how the brain works, they are gaining a better understanding of how the condition arises and how it may be treated.

ENIGMATIC ILLNESSES

MORE THAN 3,000 YEARS AGO Mursili II, king of the Hittites, was caught in a terrifying thunderstorm. The experience left him with a temporary speech impediment that went away—only to return several years later, after the monarch woke from a nightmare about the incident. His subjects attributed their king's curious ailment to the wrath of the Storm God, one of the most important deities of the ancient civilization. When modern-day scholars revisited the documents detailing the event, they interpreted it as functional aphonia (the inability to speak).

Like the Hittites, people throughout history have turned to the supernatural—gods, witchcraft and demonic possession—to explain illnesses that today would likely be diagnosed as FND. According to some historical interpretations, the first scientific attempt to account for them emerged around 400 B.C.E., when Greek physicians, including Hippocrates, coined the term "hysteria" to describe a wide collection of ailments, among them paralysis, headaches, dizziness and pain, in the belief that they were caused by the uterus (*hystera,* in Greek) wandering about the body.

Hysteria had its heyday in the 19th century, when it moved from the womb to the brain. Among several

physicians who studied it was Jean-Martin Charcot, regarded by many as the "father of neurology." At the Salpêtrière Hospital in Paris, he painstakingly detailed the symptoms of patients with hysteria and, after they died, conducted autopsies on their brains. Although Charcot was unable to identify any structural aberrations in those subjects, he was convinced that the impairments he saw were associated with unobservable, fluctuating changes in the brain, which he called "dynamic or functional lesions."

Charcot also discovered that, contrary to common belief, male hysteria was not rare; instead it often went undetected. He highlighted, for example, cases of hysteria among workers at a national railway company that had seemingly emerged after minor physical injuries. His work popularized the study of hysteria, inspiring several researchers, including Joseph Babinski, Pierre Janet and Sigmund Freud, to investigate it as well.

Unlike Charcot, however, these men viewed the condition as a disorder of the mind rather than the brain. Freud proposed that it arose when repressed trauma from childhood abuse or other disturbing events was transformed into physical symptoms; accordingly, he called it conversion disorder. That view and label cemented the displacement of the disorder from the realm of neurology to that of psychiatry and became the dogma for much of the 20th century. Perhaps coincidentally, as Freud's influence in psychiatry faded over the decades, so did cases of conversion disorder—to the point where some viewed it as a bygone ailment of the Victorian era.

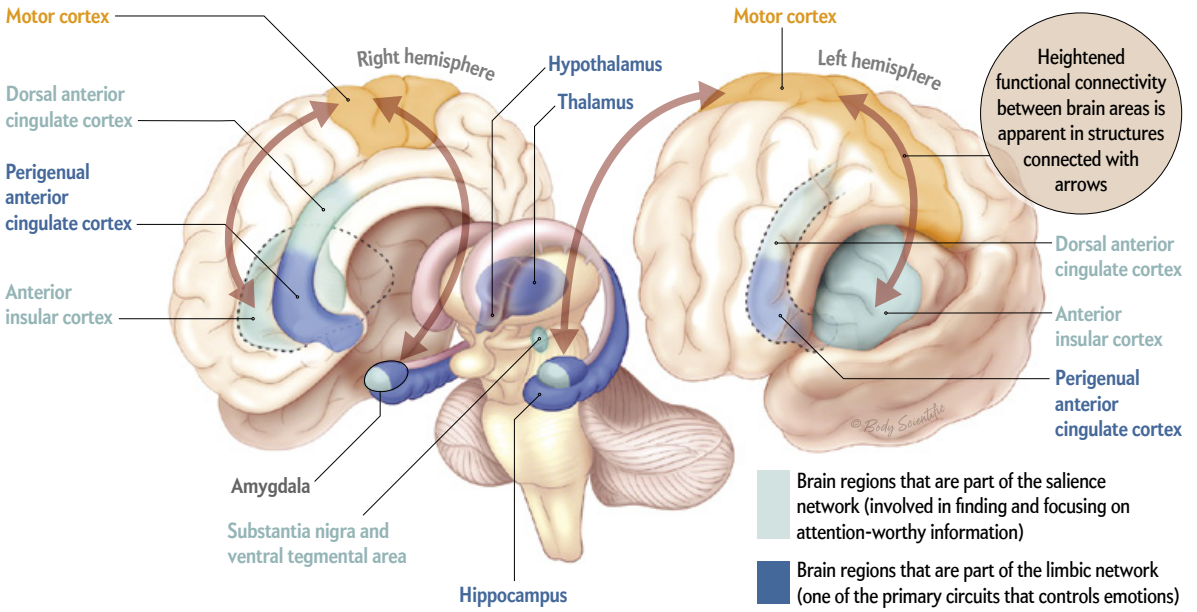
A century later a new generation of investigators has turned its attention to this condition. Careful observation of patients indicates that despite the drop in diagnoses in the latter half of the 1900s, these disorders have not disappeared. And new research reveals that the condition encompasses both neurology and psychiatry. In 2013 some physicians, concerned that the term "conversion disorder" was not widely accepted by patients and perhaps incorrectly pointed to psychology as an exclusive driver for the condition, lobbied for a change—causing FND to be included as an alternative name for the ailment in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*).

TELLTALE SIGNS

ON A COOL, SUNNY AFTERNOON in February 2020, I watched neurologist Jon Stone of the University of Edinburgh consult with first-time patients at the Anne Rowling Regenerative Neurology Clinic, where

The Brain-Body Connection

Patients with physical symptoms (such as paralysis) but no apparent injuries may have functional neurological disorder (FND). Neuroimaging has revealed subtle abnormalities in several brain regions and networks. Studies find, for example, that functional connectivity—meaning correlations in activity—is heightened between areas involved in controlling movement and regions that affect attention and emotion, as shown in the cutaway. These linkages suggest a possible mechanism for the ailment. Activity in circuits associated with a sense of agency, such as the temporoparietal junction and its connections, may also be altered.


Motor cortex
Right hemisphere
Hypothalamus
Thalamus
Dorsal anterior cingulate cortex
Perigenual anterior cingulate cortex
Anterior insular cortex
Amygdala
Substantia nigra and ventral tegmental area
Hippocampus
Motor cortex
Left hemisphere
Heightened functional connectivity between brain areas is apparent in structures connected with arrows
Dorsal anterior cingulate cortex
Anterior insular cortex
Perigenual anterior cingulate cortex
© Body Scientific
Brain regions that are part of the salience network (involved in finding and focusing on attention-worthy information)
Brain regions that are part of the limbic network (one of the primary circuits that controls emotions)

Source: "Emotional Processing in Functional Neurological Disorder: A Review, Biopsychosocial Model and Research Agenda," by Susannah Pick et al., in *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 90; 2019

Illustration by Body Scientific

his team specializes in diagnosing and treating FND. Stone listened carefully as people described when, where and how their symptoms started. He collected detailed information about their medical and personal histories and conducted a neurological examination. Then, like a detective, he pieced these details together to make a diagnosis.

In recent years FND has gone from a diagnosis of exclusion—a label doctors reserved for patients whose conditions defied all other explanations—to one made after identifying distinct signs and symptoms. These resemble those of other neurological disorders but possess identifiable differences. One example is Hoover's sign, in which weakness in a limb is temporarily corrected when the patient's attention is directed elsewhere. Another is tremor entrainment: when patients with a functional tremor in one arm are asked to start shaking the other at a regular rhythm, the affected hand will start to shake with the same rhythm as the other. This effect does not occur in people with neurodegenerative conditions such as Parkinson's disease. Clear signs of functional seizures include tightly shut eyes, rapid breathing and shaking that lasts for several minutes—features rarely seen during epileptic attacks.

CLOSE TO A THIRD OF PATIENTS REFERRED TO NEUROLOGICAL CLINICS IN SCOTLAND HAD SYMPTOMS THAT WERE MEDICALLY UNEXPLAINED.

Such indicators have been known to doctors for decades—Hoover's sign, for one, was observed by physician Charles Franklin Hoover in the 19th century. In the past, physicians would hide such signs from patients, Stone explains. But he shows them to patients to help them understand the nature of their condition and notes that physicians are increasingly taking up this practice. Seeing such clues can help a person grasp a condition that Stone likes to describe as a "software problem, not a hardware problem" in the brain.

Stone first came across these disorders as a junior doctor in the early 1990s. He found himself fascinated by them; having grown up with a stutter meant that he had experienced being unable to control his own body. And he was disturbed by how those with FND, as he prefers to call the condition, were being treated. The common attitude among medical professionals was that the symptoms were not real—at least not in the same way as those seen in multiple sclerosis or stroke, for example. Many physicians were concerned that they would either fail to identify the true cause of a patient's illness or be fooled by someone faking their symptoms. As a result, patients with FND did not receive the same level of sympathy, attention or care as those with so-called organic neurological illnesses.

Stone decided to dig deeper. During his doctoral studies at the University of Edinburgh, he met Alan Carson, who was training to become a psychiatrist and shared his interest. In 2002 the pair began to assess the scale of the problem by following the referrals to four neurology centers in Scotland over a period of 15 months. Their examination, which included more than 3,700 patients, revealed that 1,144—close to a third—had neurological symptoms deemed as partially or completely medically unexplained. Of those, only four ended up being diagnosed with another neurological issue 18 months after their initial consultation. This work demonstrated how widespread these disorders were.

Eventually Stone and Carson joined forces with Mark Hallett, a neurologist at the National Institute of Neurological Disorders and Stroke in the U.S., who had also been taking strides to advance the field. In addition to conducting his own research, Hallett had begun mobilizing a group of scientists and physicians who could contribute to the study of FND. The community gradually ballooned from several dozens of attendees at a small workshop to a full-blown society for FND, which was founded by Hallett, Carson and Stone and was inaugurated in 2019.

PREDICTIONS GONE AWRY

A YEAR AFTER LANDING in the hospital because of her FND, McNiven was referred to a psychologist. At first, she did not think she needed to be there—her symptoms had been improving with physiotherapy. After several sessions of psychotherapy, however, McNiven made a shocking discovery: she had blocked the memory of certain key events in her childhood. Among those forgotten experiences were years of physical abuse at the hands of a family friend.

Although she had initially been reluctant to consider the role of psychological factors in her illness, McNiven now says that she thinks they do contribute. "I don't think it's just purely down to that," she tells me. "But it certainly does have a big impact." Many of those who study FND today would agree—with caveats. Unlike Freud, who focused on the role of repressed trauma, researchers now recognize that myriad factors are associated with this condition. These include predisposing factors, such as adverse experiences during childhood, a previous physical injury or mood and anxiety disorders; triggers such as physical injury or a stressful life event; and maintaining factors, such as a lack of access to proper treatment or a patient's responses to and beliefs about the condition. The leading framework for thinking about FND, the so-called biopsychosocial model, takes all these factors into account.

As yet, there is no single, widely accepted explanation for how these influences come together to create

FND, but some scientists have suggested that the malady involves arguably one of the most fundamental functions of the brain: predictive processing. Championed by neuroscientist Karl Friston of University College London, predictive processing posits that the brain is constantly making and evaluating predictions by comparing the data generated from our sensory organs to internal models built from previous experience. When mismatches occur between inferences and reality, the brain either updates its existing models or sends commands back down to the body to act in ways that align with our expectations. For instance, if you want to walk, but your leg is not moving, the brain will generate a prediction error that can be resolved if you move that leg. In this way, Friston and others propose, predictions underlie everything the brain does, from perception to movement to decision-making.

Neurologist Mark Edwards of St. George's University Hospital in London and his colleagues have suggested that with FND, this predictive machinery goes awry, so that patients develop abnormal inferences of how their body should feel or function. One of the biggest drivers of this anomaly, according to Edwards, is excessive focus on one's own body. This heightened attention can be attributed to a variety of factors, including an existing physical illness, mood and anxiety disorders, or childhood abuse. When a person experiences a triggering event—say, an injury to a limb or a panic attack—this heightened attentiveness may drive one's brain to develop altered predictions about the body. In some cases, a past experience, such as exposure to sickness in the family, might also help shape these expectations.

Consider someone who falls and badly sprains a leg, resulting in a temporary loss of mobility in that limb. In most people, the brain's predictions about the injured leg's ability to move would get updated once mobility returns. This person, however, has a tendency toward mild anxiety that amplifies the levels of subconscious attention they pay to their body and has been sensitive to health events since the sudden death of a parent. These predisposing factors magnify the sensations associated with injury; in consequence, the internal model of the immobile leg persists even after the limb regains its function, leading to functional paralysis. (In some ways, this is the opposite of what happens in people who experience phantom limb syndrome. Those people are unable to update the prediction error that occurs when an expected sensation in a missing limb is not met with actual sensory feedback.)

The hypothesis that predictive processing is altered in FND patients has now been tested in a handful of experiments. In a 2014 study, for example, Edwards and his team used a task called force matching, in which a robotic device presses down

Magnus Wennman

Resignation Syndrome

Starting in the 1990s, reports of a mysterious illness began to emerge from Sweden. Children started showing up in emergency rooms in a comalike state—immobile, mute, unable to eat or drink, and unresponsive, even to pain—with no identifiable medical cause, and they remained in this state for months, sometimes years. The patients had some things in common: they were from families of refugees, most of whom belonged to ethnic or religious minority groups from former Soviet or Yugoslav states. And in many, the trigger for their illness appeared to be the rejection of an application for asylum.

Hundreds of cases of the unexplainable illness had been reported by the early 2000s. As the number of afflicted children rose, the nature of the illness became a subject of intense debate. Some opined that the children were faking or that parents were inducing the condition to obtain a residence permit—suggestions that, despite sparking outrage among both clinicians and the public, continue to circulate today. In 2014 the Swedish National Board of Health recognized it as a novel condition, *Uppgivenhetssyndrom* ("resignation syndrome"). Others said it was a manifestation of a known illness, such as severe depression, catatonia or conversion disorder.

Karl Sallin, a pediatrician at Karolinska University Hospital in Sweden, and his colleagues have proposed that the condition is a culture-bound functional neurological disorder. They suggest that resignation syndrome arises when factors such as prior psychological or physical trauma, loss of hope that asylum will be granted and fear of being deported combine with culturally specific beliefs to subconsciously prescribe how the body should respond in the face of extreme external stress. Sallin notes that the apparent specificity of this condition, which is limited to certain refugee communities in Sweden, suggests that the illness is influenced strongly by beliefs prevalent in a particular group. (Reports of a similar condition among refugees waiting for asylum in Australia on the tiny island of Nauru have also emerged, however.) As yet, experts do not agree on what these disorders are.

The number of cases of resignation syndrome has decreased since the peak in the early 2000s. Even so, hundreds of children have been diagnosed with the condition in recent years. —*D.K.*

on a finger and people are asked to match the force with their own hand. Healthy people tend to overestimate the force required by their own hand because the brain's expectations "cancel out" some of its force (a similar explanation applies to why you cannot tickle yourself). People with FND, on the other hand, were abnormally accurate, indicating that the internal prediction system was functioning differently. Even so, much more evidence is needed to prove that this mechanism provides a correct and sufficient explanation for the condition.

USING NONINVASIVE PROBES, RESEARCHERS ARE FINDING SUBTLE DIFFERENCES IN THE BRAINS OF INDIVIDUALS WITH FUNCTIONAL NEUROLOGICAL DISORDER.

PROBING THE BRAIN

LIKE CHARCOT, contemporary investigators of FND have been examining the brains of patients to find changes associated with the condition. Modern scientists, however, no longer have to wait to conduct an autopsy to peer into their subjects' skulls. Using techniques such as fMRI, researchers have begun to reveal there are indeed differences in the brains of individuals with FND. "We're beginning to identify the dynamic lesion that Charcot was looking for," says David Perez, a neurologist-psychiatrist at Massachusetts General Hospital.

With fMRI, researchers have identified distinct patterns of activity in brain areas such as the temporoparietal junction—associated with a sense of agency—in those with FND, compared with those asked to mimic the same symptoms. These findings help to confirm that unlike conditions such as factitious disorder (a severe form of which is known as Munchausen syndrome), in which patients deliberately act out other illnesses, symptoms in individuals with FND are out of their control.

Another significant discovery from neuroimaging is that people with FND have enhanced connectivity between the motor-control regions and two brain networks involved in emotional processing: the salience network, responsible for detecting and focusing on attention-worthy information, and the limbic network, one of the primary systems controlling emotion. In a 2010 study, for example, Hallett's group reported heightened linkages between the amygdala, a key region in the limbic system, and the supplementary motor area, which is responsible for preparing to initiate movements. Others, such as Perez, have shown hyperconnectivity between motor regions and salience network areas such as the insula and the anterior cingulate. These observations suggest that, at least in a subset of people with FND, the emotional circuitry might be hijacking the motor system, Perez explains.

Perez's team has also found that some risk factors may map onto these circuits. In a study published in 2020, his group reported that the magnitude of the coupling of the motor regions with the limbic and salience areas of the brain positively correlated with the degree to which patients experienced physical abuse during childhood. Perez emphasizes, however, that this will probably be relevant only to the subset of patients in whom trauma is present: in his study, a significant proportion of patients did not report any childhood physical abuse. Still, he notes that these findings point to how a risk factor such as trauma could alter brain circuits in people who develop FND.

Scientists are also investigating how factors such as stress alter brain circuits in FND. Neurologist Selma Aybek of the University of Bern says that although not all patients have a history of trauma or stress, they may possess differences in their biological stress response. Her group has found that, compared with healthy individuals, FND patients have higher levels of the stress markers cortisol and amylase and report being more stressed after taking part in a stressful task. Building on these findings, her team is using neuroimaging to examine whether there is an association between stress-related regions and agency-related regions in FND patients' brains.

Thus, a picture of the pathophysiology of FND patients is slowly emerging. But most of this work has been conducted in patients with motor symptoms, which means that sensory symptoms such as altered vision have yet to be explored. Many of these studies also have had small sample sizes, so findings will need to be validated in larger trials, says Valerie Voon, a neuropsychiatrist at the University of Cambridge, who collaborated with Hallett on several groundbreaking projects. How these neuroimaging findings fit with the predictive-processing model also remains an open question. It is plausible, Perez notes, that many of the areas identified so far may be the circuitry through which the altered predictions arise.

A BRIGHTER FUTURE

IN THE SUMMER after her second year of teacher's college in Scotland, a 19-year-old woman named Rachael Troup was rushed to the hospital with what appeared to be a stroke. Brain scans showed that she did not have a stroke, however, and tests for other neurological diseases came back normal. Eventually Troup was diagnosed with FND. But when she started treatment, it was excruciating. Neither her doctors nor her physiotherapists seemed to know much about how to treat her condition, and the exercises they made her do hurt more than they helped. "I was in pain constantly," she says.

After a few months Troup decided to stop going to physiotherapy. At the time the entire right side of her body was barely functioning, and she was using a wheelchair for mobility. After being admitted to the hospital several more times for strokelike attacks, however, Troup met Stone's team and was provided with FND-tailored care. It involved a form of physiotherapy that employs techniques such as distraction to shift the spotlight of attention away from the affected limbs while engaging in exercises to help restore normal control.

For FND patients, shifting focus away from affected limbs is often a crucial part of physiotherapy because, as Edwards's predictive-processing model suggests, attention is critical to the generation of symptoms. With attention deployed elsewhere, the brain's abnormal expectations about movement are unable to take hold. Stone and his colleagues are part of an ongoing U.K.-wide, randomized controlled clinical trial testing this type of specialized physiotherapy for functional motor disorders (a subset of FND affecting movement). In addition to retraining movement, the treatment includes educating patients about how such symptoms could arise and the physical and psychological factors that may underlie it.

To expand the tool kit of interventions for FND, researchers are also testing other alternatives. Another large clinical trial with more than 300 patients assessed the efficacy of cognitive-behavioral therapy (CBT)—a type of goal-oriented intervention focused on changing disruptive patterns of thinking or behavior—for functional seizures. The findings, published in June 2021 in *Lancet Psychiatry,* suggest that CBT may not reduce seizure frequency in all patients.

At King's College London, neuropsychiatrist Tim Nicholson and his team are examining a noninvasive method of exciting the brain known as transcranial magnetic stimulation (TMS) as a potential intervention for FND. His group completed a feasibility study, and the results were promising enough for them to initiate a larger pilot clinical trial. There are competing explanations for why TMS might work. It induces a brief muscle twitch that could kickstart the relearning of movement; stimulating brain areas altered in FND might help restore function, or it may have a placebo effect. LaFaver, working with colleagues at Northwestern University, has examined the use of meditation and mindfulness practice, which, she says, patients have anecdotally reported as helpful for maintaining the benefits of treatment.

Psychological treatments such as CBT currently remain among the first-line interventions for people with FND, according to Perez. There is a pressing need for a range of effective treatments, however: the prognosis remains poor. It is still relatively uncommon for FND patients to completely regain function, and relapses occur often. According to a 2014 meta-analysis of 24 studies, on average 40 percent of patients reported similar or worsened symptoms seven years after their initial diagnosis. On top of that, patients still experience high levels of stigma and have trouble accessing treatment, LaFaver says. "I still think we have a long way to go."

Fortunately, the situation is changing. As researchers' interest in FND surged over the past decade, so did the number of FND clinics around the world. Patients are speaking up as well. In 2012, for example, the international charity FND Hope was established with the aim of raising awareness and empowering patients.

Still, debates linger—and are reflected in an ongoing tussle over the name of the illness. In what fraction of patients are psychological factors present, for example? Do symptoms primarily arise through conversion of stressors, or are other explanations also necessary? One meta-analysis found that reports of stressors in FND patients vary between 23 and 86 percent in different studies. W. Curt LaFrance, Jr., a neurologist-psychiatrist at Brown University, says that it takes more time to identify such stressors than some doctors can devote to an individual patient—which may account for this enormous spread. In his clinic and in the scientific literature, he has consistently seen evidence of psychological stressors having converted into physical symptoms, and, accordingly, he supports use of the term "conversion disorder." Partially reflecting such views, the older name remained when FND was added to the *DSM-5,* but the need to identify psychological factors for a diagnosis was dropped. That excision also remains contentious.

One thing is clear, however: because the condition lies at the crossroads of neurology and psychiatry, insights from both fields will be necessary to solve the puzzle. This blurring of the line between mental and physical illness is a growing trend. Scientists now understand, for example, that stress—a psychological factor—can predispose people to Alzheimer's disease and that inflammation—a physical factor—may give rise to depression. In addition, traditional neurological diseases such as epilepsy and stroke are often associated with mood and behavioral disturbances. "The brain doesn't separate into neurology and psychiatry," Perez says. "We need a new science of brain and mind that really encapsulates that brain health equals mental health and physical health."

Over the past few years McNiven has frequently used a wheelchair because of her FND. But with the help of both physiotherapy and psychotherapy, she has started to recover. Her symptoms are not gone—she still experiences a lack of sensation in her legs, altered vision and pain—and some days are worse than others. "I constantly feel like I'm fighting against my body," McNiven says. But she hopes to make a full recovery. "You've got to keep that positive attitude to keep fighting through it. There's always hope you can get there with this condition." SA

Diana Kwon is a freelance journalist who covers health and the life sciences. She is based in Berlin.

THE UNDISCOVERED ILLNESS

Hundreds of thousands of people experience mania without ever getting depressed. Why does psychiatry insist on calling them bipolar?

By Simon Makin

Photography by The Voorhes

IN OCTOBER 1997, at the age of 58, David Ho had an unusual experience while listening to a recording of Bach. "I began to dance and pretended to conduct," he says. "And as I practiced, instead of following the music, I felt as if I were creating it. I entered into a state of selfless oblivion, like a trance. My mind exploded. Flashes of insight rained down, and I saw beauty everywhere, in faces, living things and the cosmos. I became disinhibited, spontaneous, liberated."

Ho was in the grips of his first episode of mania. His description sounds like an enviable burst of creative energy, but the symptoms of mania can also include inflated self-esteem, grandiosity, racing thoughts, extreme talkativeness, decreased need for sleep, increased activity or agitation, reckless behavior, delusions and other psychotic events. Severe episodes can impair day-to-day functions, sometimes enough to require hospitalization.

Perhaps the most surprising thing about such cases is that in the eyes of the psychiatric profession, mania does not exist as a distinct and unalloyed condition. Mania is usually known as the upside of bipolar disorder. For most people, it occurs alongside periods of depression, the downside. But Ho, who has had at least 20 manic episodes since 1997, has never suffered from depression. Thousands of people in the U.S. share that experience. Unlike those who experience only depression, however, patients with mania alone are lumped with those who have bipolar disorder. This puts psychiatry in the strange position of claiming that depression by itself is different from depression accompanied by mania but that mania by itself is not.

Most psychiatrists agree unipolar mania exists, but there is debate about whether it differs sufficiently from bipolar disorder in important enough ways to warrant a distinct diagnosis. Central to that debate is the tension in psychiatry between fewer, broader categories and more numerous, tightly defined ones. But the missing diagnosis may have consequences for patients: some studies suggest that people with unipolar mania may respond differently to certain treatments. If, as some researchers believe, unipolar mania and bipolar disorder differ in their underlying biology, classifying mania separately could speed the development of new treatments that are more personalized and effective. But because unipolar mania is far less common than bipolar disorder, research into the condition has been both scant and equivocal.

As both a patient and a clinical psychologist, Ho is well placed to advance this debate. In 2016 he published a self-study in the journal *Psychosis* cataloguing his symptoms, which include enhanced recall, increased empathy and spiritual experiences. He has suffered some ill effects, including severe fatigue, confusion and behavior that caused concern among friends and colleagues: he once burst into tears while delivering a lecture. But his professional training has helped him control his impulses and avoid delusional thinking. On balance, he believes that his madness, as he calls it, has enriched rather than damaged his life. "I'm aware my case may be atypical," Ho says. "Precisely for this reason, it challenges prevailing psychiatric beliefs that fail to acknowledge the positive value of mental disorders."

A MODERN ILLNESS

CREDIT FOR THE MODERN CONCEPT of bipolar disorder usually goes to 19th-century French psychiatrist Jean-Pierre Falret, who called it *folie circulaire,* or "circular insanity," for its periods of pathologically elevated and depressed moods, usually separated by symptom-free periods of varying length. This idea became gospel in the early 20th century, when a father of modern psychiatry, Emil Kraepelin, proposed a historically significant hypothesis.

At the time, psychiatry drew a distinction between so-called reactive psychoses, which were seen as a response to outside events, and endogenous psychoses, which were innate. Kraepelin divided all endogenous psychoses into two broad classes: dementia praecox—now known as schizophrenia—and manic-depressive insanity, now known as bipolar disorder. Endogenous depression was therefore classed as a form of manic-depressive insanity. All mania also fell under the same rubric because mania was thought never to be a reaction to outside events. There were dissenters, notably the renowned German neurologist Carl Wernicke, who held that mania was related to hyperactivity of neural firing and depression to decreased neural activity. But Kraepelin's idea dominated and persists in today's diagnostic system.

The question of what to include under the umbrella of bipolar disorder reignited in 1966. In separate investigations, psychiatrists Carlo Perris of Umeå University in Sweden and Jules Angst of the University of Zurich in Switzerland each studied some 300 patients with either true bipolar disorder or depression alone and more than 2,000 of their close relatives.

Both researchers found that relatives of the bipolar patients had more mood disorders than those of patients with depression alone. They also discovered that although bipolar illness was common in the relatives of bipolar patients, it was no more common in relatives of depressed patients than in the general population. These findings, Perris and Angst argued, suggested that bipolar disorder and depression were genetically different conditions.

As a consequence, when the third edition of the *Diagnostic and Statistical Manual of Mental Disorders,* or *DSM,* appeared in 1980, it included major depressive disorder as a condition distinct from bipolar disorder. Perris and Angst's studies focused only on depression and did not address mania. "There weren't enough cases of pure mania to do anything reasonable," Angst says.

Whether unipolar mania should have its own diagnosis is complicated by bipolar disorder's clinical diversity. The manic

and depressive phases vary in severity and the extent that one or the other dominates. The pattern of episodes varies unpredictably and from patient to patient. Mixed states, involving aspects of opposite mood extremes simultaneously, sometimes occur, too. Indeed, many psychiatrists argue that mood disorders are best thought of as lying on a spectrum, ranging from major depression through various bipolar presentations to pure mania.

IN SEARCH OF A SUBTYPE

THE VARIABILITY OF SYMPTOMS, along with findings from large psychiatric genetics studies that implicate numerous biological factors, suggests that bipolar disorder includes a range of subtly different conditions. "One reason we still have limited understanding of bipolar disorder after 50 years of intense research is that it's treated as one entity, and it's clearly not," says psychiatrist Paul Grof of the University of Toronto.

The resistance to subtyping may be the result in part of changes in research funding over the past few decades, as the pharmaceutical industry has taken over progressively more psychiatric research from universities, Grof says. Drug companies generally just want to know if a new drug is better than a placebo, and the larger the patient group, the greater the likelihood of finding a significant difference. Subdividing bipolar disorder into smaller populations would complicate these efforts. The industry also prefers to study diagnoses recognized by the Food and Drug Administration—and unipolar mania is not on its list.

MANIA IS USUALLY KNOWN AS THE UPSIDE OF BIPOLAR DISORDER. FOR MOST PEOPLE, IT OCCURS ALONGSIDE PERIODS OF DEPRESSION, THE DOWNSIDE.

Institutional inertia can also come into play. Every rewrite of the *Diagnostic and Statistical Manual of Mental Disorders* is a laborious process. Each edition is based on the previous one, and any change must be backed by fresh evidence, with papers submitted to committees justifying the decision. The last edition, *DSM-5,* was published in 2013, and in the view of the committee tasked with reviewing mood disorders, unipolar mania was covered by the bipolar diagnosis known as BP-I, which is mania with or without associated depression. "There was very limited discussion as to whether mania should be separate because the onset and course of illness weren't seen as that different from BP-I," says psychiatrist Trisha Suppes of Stanford University, who was a member of the *DSM-5* work group for mood disorders.

The lack of a separate diagnosis may be making evidence harder to gather. The standardized clinical interview used under the *DSM* to make diagnoses for research studies has no category for unipolar mania, meaning investigations of the condition would have to rely on ad hoc techniques that might not align with those used in other studies. Unipolar mania is thus at the hub of a catch-22: the absence of a diagnosis is an impediment to research, and the paucity of research makes the creation of a diagnosis less likely.

In studies that do occur, the lack of a formal designation for unipolar mania makes it difficult to compare results. "A major problem is definitions," says Allan Young, a psychiatrist at King's College London. One source of disagreement is the severity of symptoms necessary for a case to qualify as mania. Another is the frequency of episodes. Some studies include anybody who has had at least one episode of mania with no history of depression, whereas others require three or four. Still others stipulate a minimum number of years of illness. These differences have led to widely disparate prevalence estimates for unipolar mania, ranging from 1.1 to 65.3 percent of patients with bipolar disorder.

Most of the studies completed so far also have methodological problems. The bulk are retrospective, in which researchers simply ask participants to recount past experiences—a process known to underestimate depression, perhaps inflating estimates of pure mania. Prospective studies that follow patients for years and include periodic assessments are better. "What you really want is someone who's lived their whole life, had multiple episodes of mania, and never had depression," Young says. "The first lady I saw like this died in her late 60s and had her first episode at 21, which is getting on for 50 years, so that's very convincing."

One of the longest prospective studies, led by David Solomon, now professor emeritus at Brown University, began in 1978 and was published in 2003. It began as a study of 229 bipolar patients, 27 of whom had mania with no history of depression. The investigators followed those 27 patients for up to 20 years; seven of them remained free of depression throughout the period. The results suggest that of the original 229 patients, 3 percent had unipolar mania. Solomon does not advocate the creation of a separate diagnosis for unipolar mania unless future research establishes differences in genesis, prognosis or treatment response. But if the rate reported in the study held for the general population, the number of people with unipolar mania in the U.S. would be around 100,000—and there would be hundreds of thousands more worldwide.

The stories of people with unipolar mania help to explain why some researchers are convinced that the disorder is a separate entity. Lindsey, a ski coach from Portland, Me., is one such case. She was 18 when she had her first experience of mania. Eighteen years later she has never been depressed, yet she still has a diagnosis of bipolar disorder. "I'm the happiest person I know," she says. "I never accepted my diagnosis." As a result, she rejected treatment and continued to have episodes. She has been hospitalized five times and has landed in jail more than once.

Lindsey's episodes start with euphoria but can spiral into delusions and difficulty speaking. While manic, she feels no fatigue, hunger or pain. One such episode, in her late 20s, began on a hike in New Mexico when she was overcome by a vision that the world was coming to an end. Lindsey called her father, who flew out to meet her and drive her home to Maine. "She had medication," her father says. "She just wasn't taking it." Early in the morning on an overnight stop in Nashville, Lindsey started playing the piano in the hotel lobby. An employee called the police, and Lindsey fled in the car.

In the adventure that followed, she deliberately got lost, buried her possessions near a railroad track and abandoned the car.

She then hopped a freight train, got off in the middle of rural Tennessee, climbed out of a rock-walled valley and wandered into a chapel, where the pastor was able to glean enough information to contact her family. Shortly after resuming the drive home, Lindsey ran away from her father at a highway rest stop and started picking daisies in a fenced-off electrical area. The police were called again, and although the officer urged her to leave with her father, she insisted on being arrested.

In her cell, a guard pepper-sprayed her, and she ended up in the office of the jail's counselor. Lindsey was barely able to speak at this point, but she wrote "unipolar" repeatedly on a blackboard. The counselor then read Lindsey a description of mania. She credits this encounter as the moment she accepted the need to take medication. The counselor gave her Zyprexa (olanzapine), an antipsychotic. She recovered and takes it to this day, though not without reservations. "My medication is like a dose of sadness, hunger, fatigue and pain," she says. Lindsey was euphoric throughout her ordeal, even while being pepper-sprayed. Only the people around her suffer. "I feel like I've been blessed with this illness that makes me so happy," she says, "but I feel selfish because of how it affects my family."

Lindsey married Andy, a journalist, in 2015, not long after he witnessed her last hospitalization. "It made the relationship stronger in the end," he says. "I got to see her as she clawed her way back to sanity. It was impressive." The most important factor in her treatment is whether a physician accepts that she is not bipolar. "When that's ignored, she no longer trusts that person," Andy says.

Psychic Fuel for the Creative Brain

The mad genius may be more than a cliché

Of all the tropes of artists and mental afflictions, the most enduring is the one of a genius in the throes of mania. Iconic figures ranging from William Blake to Ernest Hemingway to Kurt Cobain were known or believed to have bipolar disorder. The association has intuitive appeal: the euphoria, abundant energy and racing thoughts of mania are credible fuel for creativity.

Scientific evidence for the association has mostly been inconclusive. Much of the data comes from historical sources, and most accounts are anecdotal. Modern investigative techniques have revealed surprisingly little about what happens in the brain during mania, partly because brain imaging requires minimal head movement, so scanning someone in a floridly manic state is a challenge. As a dynamic process involving the interplay of multiple brain networks, creativity is also difficult to research.

But comparing findings from research into bipolar disorder with certain studies of creativity reveals hints of a link: cognitive "disinhibition" seems to be a feature of both the creative state described as being in the "flow" and altered brain circuits in bipolar disorder.

Brain-imaging studies have found reduced activity in a part of the prefrontal cortex that helps to regulate emotion, which may be linked to impaired impulse control and extremes of mood in people with bipolar disorder. (The prefrontal cortex is the brain's "orchestra conductor" responsible for directing various mental processes.) Some of these studies have also found diminished activity in an area involved in suppressing the kind of spontaneous thought that appears to well up from the unconscious, seemingly out of nowhere.

These results are reminiscent of a 2008 study of improvising jazz musicians and a 2012 study of freestyling rappers, conducted by the team of speech neuroscientist Allen Braun, then at the National Institutes of Health, which found reduced activity in the part of the prefrontal cortex that inhibits spontaneous cognition. The researchers also found an increase in activity in a section of the prefrontal cortex that is part of the so-called default mode network, which revs up when a person is not focusing on a task but rather

IT ALL GETS REAL

A CURIOUS QUIRK in the tale of this neglected disorder is that prevalence estimates vary worldwide and are consistently higher in non-Western countries. After qualifying in South Africa in 1997, psychiatrist Christoffel Grobler worked in an inpatient unit in Ireland, where his bipolar patients were mostly in depressed states. When Grobler returned to South Africa in 2009, he noticed the opposite pattern: his patients were mostly in manic states. To investigate, in 2010 he and his colleagues interviewed 103 bipolar patients in three hospitals, using a standard diagnostic questionnaire. They found that 32 percent of patients qualified as unipolar, defined as having at least five manic episodes over four or more years. "When I present this at conferences, people come up and say, 'We see this all the time,'" Grobler says.

Regional variations are tricky to interpret because cultural differences come into play: depression is more likely to be considered part of normal life in Africa, for example. The quality and procedures of health-care systems differ, and other causes, including infection or intoxication, may be a factor. But Grobler is convinced the geographical differences are genetic in origin and that unipolar mania therefore represents a distinct condition.

Getting to the bottom of this question will require large, multicultural international studies. In the meantime, scientists are trying to compensate for a shortage of data. One reason most early studies failed to find differences between mania and bipolar disorder may be that they are so slight that they can be reliably detected only in large samples. Now in his 90s, Angst addressed this problem by consolidating data from nine epidemiological studies conducted in the U.S., Germany, Switzerland, Brazil and Holland. That study, published online in November 2018 in *Bipolar Disorders,* found that people with unipolar mania were more likely to be male but less likely to have attempted suicide or to suffer from anxiety, drug use and eating disorders. Angst and his colleagues claim these findings suggest unipolar mania "should be established as a separate diagnosis."

Some of these findings align with three reviews of research on unipolar mania published in the past seven years. All three found that unipolar mania is less likely to co-occur with anxiety (which often accompanies depression) but more likely to come

is imagining things or ruminating on the past. They believe what they observed reflects relaxation of focused attention and control, making way for a creative thought process in which inspiration bubbles up from the unconscious. Other studies have found reduced thickness of certain cortical regions in both creative and bipolar brains, which may be linked to altered brain activity and disinhibited cognition.

Another element in the thinking patterns of creative and manic people is the ability to make mental connections that elude others. Neuroscientist Nancy Andreasen of the University of Iowa has found that creative people show greater activity in the so-called association cortices, which are regions tasked with linking related elements of cognition. These brain areas are not devoted to processing specific sensory or motor functions but instead engage with tasks such as tying together a written word with its sound and meaning. Andreasen believes creative ideas probably happen when these types of associations occur freely in the brain during unconscious mental states, when thoughts become momentarily disorganized—not unlike psychotic states of mania.

This observation resonates with clinical psychologist David Ho, who has experienced racing thoughts and extraordinarily enhanced recall during manic episodes, letting him write without inhibition or self-doubt. "With repression vanished, my mind functioned with holistic oneness," he says. "Creative ideas rained down faster than I could cope." Researchers do not know if the association cortices are more active in mania, but all these findings suggest that at key moments of the creative process, our thought processes flow more freely, with novel combinations of sights, sounds, memories, meanings and feelings producing insight and originality in creative work akin perhaps to what happens during mania.

Of course, mental illness is neither necessary nor sufficient for creative talent, and severe manic episodes most likely are too debilitating for any kind of sustained activity. But researchers have found that family members of people with bipolar disorder also tend to be more creative than average, supporting the idea that mild manifestations of the disorder may furnish cognitive benefits.

It is important not to romanticize conditions that mainly cause suffering, but evidence that mania can enhance creativity in some people may help reduce the stigma of a diagnosis. "It is possible to retain a measure of madness in dignified living," Ho says, "and of dignity even in a state of madness." —*S.M.*

with psychotic symptoms. Unipolar mania also seems to confer less social impairment and involve fewer recurrences and better remission rates than bipolar disorder.

Perhaps most important, people with unipolar mania show subtle differences in their response to drugs administered as part of preventive treatment. Three studies found that patients with unipolar mania respond less well to lithium (a mood stabilizer and first option for bipolar) than those with true bipolar disorder do.

One of these studies, published in 2012 by Olcay Yazici and Sibel Çakir, then both at Istanbul University, also examined the question of whether unipolar mania is merely bipolar disorder weighted to the manic end of the spectrum—so-called dominant manic polarity. They divided 121 patients into two groups, 34 with unipolar mania and 87 with classic bipolar disorder. As the earlier studies found, the unipolar group had a lower response rate to lithium, and their response to another frontline bipolar treatment, the anticonvulsant Depakote (divalproex sodium), was no different.

The researchers next grouped all 121 patients according to whether the majority of their episodes were manic or depressive and then created a further division of patients whose manic episodes accounted for at least 80 percent of the total. A smaller percentage of patients who had at least a majority of manic episodes responded to lithium than among patients who had more depressive episodes, and this difference was greater for patients whose mania put them in the 80 percent group. Most telling, when those with unipolar mania were excluded from this analysis, these differences disappeared, suggesting the treatment difference relates to unipolar mania and not to dominant manic polarity and thus implying that unipolar mania is its own entity.

THE WAY AHEAD

THOSE WHO ARE OPPOSED TO a separate diagnosis sometimes point out that the mania in unipolar mania is indistinguishable from that in bipolar disorder. But the same is true of depression, and many studies have found differences in the brains of people with major depression compared with those of individuals with bipolar disorder. Future work that compares brains of people with unipolar mania and bipolar disorder might be just as revealing.

Biological and brain-imaging studies of unipolar mania are rare. But one from several decades ago gives clues to differences in physiology. A 1992 CT scan study led by Sukdeb Mukherjee of the Medical College of Georgia at Augusta University found that unipolar mania patients had smaller third ventricles (one of four interconnected cavities in the brain that let cerebrospinal fluid flow) than bipolar patients did.

This result is intriguing because subsequent studies found that bipolar patients who experienced multiple episodes have larger ventricles than people who are experiencing their first episode or healthy control subjects, a hint that enlarged ventricles may be linked to pathology. The implication that unipolar mania may not cause as much damage in the brain tallies with the better outcomes associated with the condition.

Creating a separate diagnosis for unipolar mania remains controversial. An interim step would be to recognize it as an official subtype of bipolar disorder. Such a move might encourage research and raise awareness among clinicians. "There's a mystery here we don't understand: Why do some people get mania and then depression, whereas others stay unipolar manic?" Suppes asks. "It's deserving of more research than it's gotten so far." Further investigation might also benefit patients who do not identify with other labels. Lindsey pleads, "The most important thing my doctor could do for me is say, 'I'm sorry, we were wrong—you're not bipolar, you're unipolar.'" SA

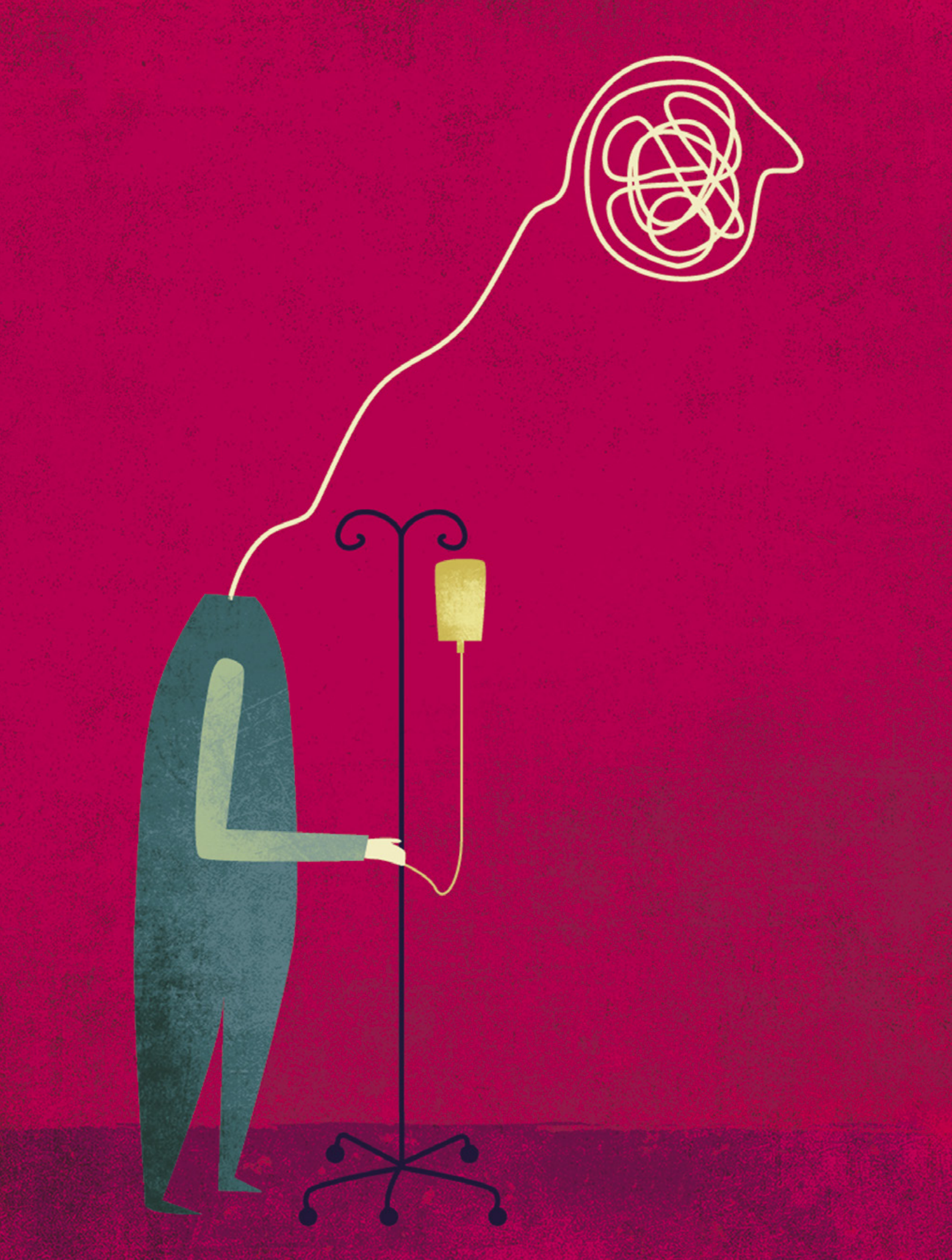
Simon Makin is a freelance science writer based in London.

COULD COVID DELIRIUM BRING ON DEMENTIA?

Delirium is very common on COVID wards. Researchers are testing whether these temporary bouts of confusion could bring on permanent cognitive decline

By Carrie Arnold

Illustration by Fatinha Ramos



IN HER JOB AS A PHYSICIAN AT THE BOSTON MEDICAL CENTER, Sondra Crosby treated some of the first people in her region to get COVID. So when she began feeling sick in April 2020, Crosby wasn't surprised to learn that she, too, had been infected. At first her symptoms felt like those of a bad cold, but by the next day she was too sick to get out of bed. She struggled to eat and depended on her husband to bring her sports drinks and fever-reducing medicine. Then she lost track of time completely.

For five days Crosby lay in a confused haze, unable to remember the simplest things, such as how to turn on her phone or what her address was. She began hallucinating, seeing lizards on her walls and smelling a repugnant reptilian odor. Only later did Crosby realize that she had had delirium, the formal medical term for her abrupt, severe disorientation.

"I didn't really start processing it until later when I started to come out of it," she says. "I didn't have the presence of mind to think that I was anything more than just sick and dehydrated."

Physicians treating people hospitalized with COVID report that a large number experience delirium and that the condition disproportionately affects older adults. An Aparil 2020 study in Strasbourg, France, found that 65 percent of people who were severely ill with the novel coronavirus had acute confusion—a symptom of delirium. Data presented at the 2020 meeting of the American College of Chest Physicians by scientists at the Vanderbilt University Medical Center showed that 55 percent of the 2,000 people they tracked who were treated for COVID in intensive care units (ICUs) around the world had developed delirium. These numbers are much higher than doctors are used to: usually about one third of people who are critically ill develop delirium, according to a 2015 meta-analysis [see "How Common Is Delirium?" in box on page 121].

Delirium is so common in COVID that some researchers have proposed making the condition one of the disease's diagnostic criteria. The pandemic has sparked physicians' interest in the condition, says Sharon Inouye, a geriatrician at the Marcus Institute for Aging and Harvard Medical School, who has studied delirium for more than 30 years.

As clinicians face the immediate realities of confusion and agitation on their wards, Inouye and other researchers are concerned about the future. In the past decade long-term studies have revealed that a single episode of delirium can increase the risk of developing dementia years later and accelerate rates of cognitive decline in those who already have the condition. The reverse is also true: having dementia makes someone more likely to develop delirium. A set of simple steps, such as ensuring a family member is present to help people orient themselves, can reduce the incidence of delirium by 40 percent, but doctors struggle to follow that advice in COVID wards.

But the links between delirium and dementia have been difficult to untangle: researchers need to follow patients for years to get results. The surge in people with delirium produced by the pandemic has focused attention on the condition and provided scientists with a unique opportunity to follow patients

Morteza Nikoubazl/NurPhoto/Getty Images

VISITS FROM relatives are a source of comfort for people with delirium, a common symptom of COVID, but many hospitals had strict no-visitor policies at the start of the pandemic.

and determine whether and how delirium might affect long-term cognition. Researchers have launched several studies to explore the long-term neurocognitive impacts of COVID, including dementia, and Inouye and others hope that this work will allow researchers to explore the links between the two conditions in real time.

If the pandemic can be said to have a silver lining, Inouye says, it has been to spur interest in how delirium can lead to dementia—and vice versa. What is more, says Catherine Price, a neuropsychologist at the University of Florida, the spread of COVID "has highlighted the blurring of the lines between delirium and dementia, especially with more older adults in our populace."

NEGLECTED CONDITION

INOUYE'S INTEREST in delirium began when she landed her first job as an internal medicine physician at a Veterans Administration hospital in Connecticut in 1985. In her first month there, she treated more than 40 people for a variety of conditions. Six of them developed delirium during their stay; none seemed to return to their previous level of physical and mental health. To Inouye, the connection between her patients' delirium and their poor prognosis was obvious. When she confessed her suspicions to her bosses, however, they just shrugged. Their attitude, Inouye says, was that delirium was just one of those things that happened.

"Why is it okay for older adults to come in the hospital and lose their minds?" Inouye asked. Answering this question, she says, would be "an uphill battle my entire career."

Shortly after, she began a two-year fellowship to study the condition in depth. Her work showed that delirium occurs when several stressors converge. Preexisting vulnerabilities such as chronic disease or cognitive impairment can combine with precipitating factors, including surgery, anesthesia or overwhelming infection, to cause a sudden onset of confusion, disorientation and attention difficulties, especially in older adults.

"Delirium easily occurs when the brain is unable to compensate for a stressful situation," explains Tino Emanuele Poloni, a neurologist at the Golgi Cenci Foundation near Milan, Italy. Researchers think that the underlying biological causes are inflammation and

an imbalance in neurotransmitters—chemical messengers such as dopamine and acetylcholine.

Inouye's mounting clinical experience has taught her that regardless of what precipitates delirium, around 70 percent of those with symptoms eventually recover completely. In the 30 percent who do not, however, an episode of delirium predicts a downward spiral over a period of months that leads to profound cognitive impairment or even to symptoms of dementia.

More formal studies have reinforced the link, to varying degrees. Inouye investigated a group of 560 people aged 70 or older who had undergone surgery, and she saw that cognitive decline over the subsequent 36 months was three times faster in those who developed delirium than in those who did not have the condition. A 2020 meta-analysis of 23 studies showed that delirium during a hospital stay was associated with 2.3 times greater odds of developing dementia. And work by a team of Brazilian scientists showed that in a group of 309 people with an average age of 78 years, 32 percent of those who developed delirium in the hospital progressed to having dementia, compared with just 16 percent of those who did not become delirious [see "Delirium and Cognitive Decline" in box on opposite page].

"WE DON'T UNDERSTAND THE MECHANISMS OF DELIRIUM AT ALL—WE REALLY DON'T. AND THERE IS NO SUCCESSFUL MANAGEMENT OF DELIRIUM FROM A PHARMACEUTICAL STANDPOINT."
—CATHERINE PRICE *UNIVERSITY OF FLORIDA*

What is more, the longer a person is delirious, the greater their risk of subsequent cognitive impairment, according to a 2013 study by psychologist James Jackson of Vanderbilt and his colleagues. Work by Inouye, Jackson and other researchers found that the reverse was also true: even after controlling for age, existing dementia symptoms increased the chances of developing delirium.

CAUSING CONFUSION

SCIENTISTS STILL DO NOT AGREE ON whether the link between delirium and dementia is strong only in those who would have developed dementia anyway or whether delirium increases the risk of cognitive decline even in individuals who are not predisposed to it. Nor can they say precisely what it is about delirium that may provoke dementia. If researchers could identify these connections, then perhaps they could prevent delirium from escalating into dementia.

"We don't understand the mechanisms of delirium at all—we really don't. And there is no successful management of delirium from a pharmaceutical standpoint," Price says.

Scientists have developed three hypotheses to explain how delirium might provoke dementia. One line of thinking holds that an accumulation of toxic cellular trash in the brain could cause short-term delirium and lead to longer-term damage. The body usually clears this molecular rubbish by way of the bloodstream and the glymphatic system, which is a network of channels filled with cerebrospinal fluid. Damage to vessels from an acute episode of delirium could persist and trigger dementia, or a brain that experiences delirium could become more prone to vascular problems in the future.

The second suspect is inflammation, which often troubles people who are hospitalized for infections, respiratory distress or cardiovascular disease. Surgery and severe infections can cause a buildup of cellular detritus in the brain, which triggers more inflammation. This short-term, all-hands-on-deck reaction safeguards the brain because it clears the harmful debris, and the inflammation ultimately dies down. That is not the case for those who develop delirium, Inouye says. Persistent inflammation can trigger an acute episode of delirium and cause neurons and associated cells, such as astrocytes and microglia, to deteriorate, leading to cognitive damage.

The third idea is what is referred to as the threshold hypothesis. Someone with dementia (even in the earliest stages) has fewer connections between neurons and can show damage to the insulation that wraps them and helps to convey signals, known as white matter. This loss strips the neurological reserves that help the person to cope with inflammation or infection, throwing them over the edge not just into delirium but into a more advanced dementia.

Even though the genesis of delirium and its molecular connections to dementia remain unknown, Inouye has managed to find a way to cut rates of delirium in hospitals. She created a program of simple strategies known as HELP (Hospital Elder Life Program), which focus on reducing sedation even during mechanical ventilation, paying close attention to nutrition and hydration, and ensuring the presence of family members to help reassure and orient patients. A 2015 meta-analysis showed that these steps reduced delirium by around 40 percent. Hospitals around the U.S. began instituting these simple protocols. Then COVID struck and made them all but impossible.

DEMENTIA SURGE

AS CROSBY ENDURED coronavirus-induced delirium in her Boston bedroom, Poloni was treating delirious

people with COVID in Lombardy—Italy's ground zero for the coronavirus. Many of Poloni's patients already had dementia, and like many physicians, he was watching for common symptoms of respiratory infections such as fever, cough and difficulty breathing. But some of his patients did not show those signs at all. Instead they mostly became "dull and sleepy," Poloni says. Others became restless and agitated—all signs of delirium. It was so prominent that Poloni argued that delirium should be added to the virus's diagnostic criteria. Inouye has made that argument, too, and it is supported by a study she published in October 2020 showing that 28 percent of older adults with COVID have delirium when they present to the emergency department.

The high numbers of people who developed delirium immediately made Inouye, Price and other researchers worry that the pandemic could lead to a surge in dementia cases in the coming decades, on top of the increase in cases as a result of aging populations [see "The Cost of Delirium?" in box on this page]. "Is there going to be an increase in dementia from people who had COVID during adulthood or midlife?" asks Natalie Tronson, a neuropsychologist at the University of Michigan. "What happens over the next decades as the population ages more?"

In a concerted effort to find answers, institutes around the world have funded a variety of studies into the long-term cognitive effects of COVID, some of which look at delirium. One such study in the U.S. tracked people who were treated in the hospital for COVID, many of whom developed delirium during their stay, and measured their cognitive and psychiatric function. An international study is planned to measure the prevalence of delirium in patients with COVID in ICUs, as well as identifying factors that predict long-term outcomes.

A separate study in Germany and the U.K. is also tracking neurocognitive outcomes in people with COVID to determine how delirium affects brain function months later. Another research project led by a team at Vanderbilt is looking for an alternative to commonly used sedatives such as benzodiazepines, which are known to increase delirium. The researchers are testing a sedative called dexmedetomidine to evaluate whether it is a safer option for people who are hospitalized with COVID.

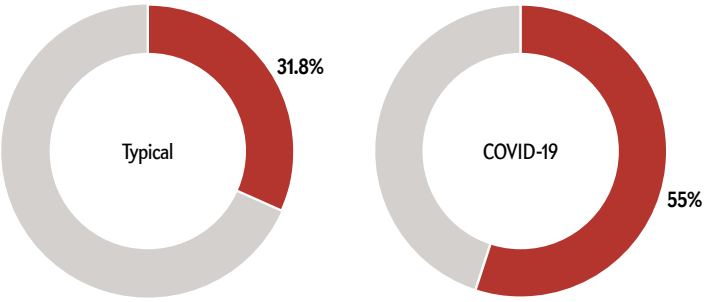
Inouye and Tronson hope that the funding of these long-term studies will lead to ongoing scientific interest in the delirium-dementia connection—and provide some insight.

"It's going to be, I think, a little bit frightening and a little bit enlightening about how illness affects dementia risk but also about what other lifestyle and genetic protective factors can influence risk as well," Tronson says. "We're learning quickly, but there are still a lot of black boxes." SA

Carrie Arnold is an independent public health reporter based in Virginia.

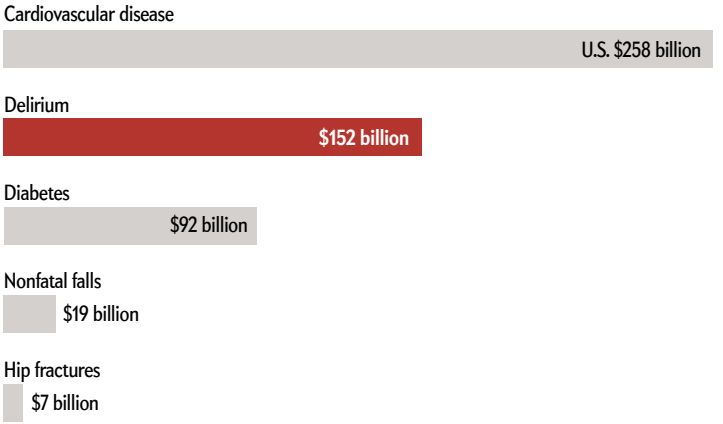
How Common Is Delirium?

Typically one third of people who are critically ill will have an episode of delirium; for COVID-19, the proportion rises to more than half.



The Cost of Delirium?

In the U.S., the annual health-care costs for delirium are higher than those for many other conditions common in older adults.



Delirium and Cognitive Decline

People who experience delirium during hospitalization are at increased risk of cognitive decline after discharge, according to a study of 309 patients in Brazil.

Nature; Sources: "Outcome of Delirium in Critically Ill Patients: Systematic Review and Meta-analysis," by J.I.F. Salluh et al., in *BMJ*, Vol. 350; 2015 ("How Common Is Delirium?"); "Delirium and Post-Discharge Dementia: Results from a Cohort of Older Adults without Baseline Cognitive Impairment," by F. B. Garcez et al., in *Age and Ageing*, Vol. 48, No. 6; November 2019 ("The Cost of Delirium?"); "The Importance of Delirium: Economic and Societal Costs," by D. L. Leslie and S. K. Inouye, in *Journal of the American Geriatrics Society*, Vol. 59, Suppl. 2; November 2011 ("Delirium and Cognitive Decline")

OPINION

A Psychedelic Renaissance

Psilocybin and MDMA represent a first wave of therapies for psychiatric disorders that help patients by changing the way they view reality

By Danielle Schlosser and Thomas R. Insel

PSILOCYBIN MUSHROOMS

Roger Cremers/Bloomberg/Getty Images

THROUGH A MAY 2021 ARTICLE HE wrote in the *Independent,* we learned about Steve Shorney, who lived with depression for most of his life despite years of psychotherapy, medication, yoga and many other attempts at holistic treatments. With his decision to enroll in a psilocybin clinical trial at Imperial College London, his life "radically changed." Psilocybin was different from every other treatment or experience he had had. As he recalled in the *Independent* article, "I had seen an alternative reality, another way of being, and knew beyond anything I'd known before that day that life is extraordinary. And in that moment I felt happier, more alive, and more Me than I imagined was possible."

The use of psychedelics, especially psilocybin and MDMA (also known as Ecstasy or Molly), is undergoing a renaissance. More than five decades ago psychedelics were an active area of research, with more than 40,000 patients receiving LSD or psilocybin for alcoholism, anxiety or depression. While we do not have rigorous clinical trials from that time, the use of these drugs garnered both scientific and public interest, with Hollywood celebrities such as Cary Grant promoting their use. But the War on Drugs, beginning during the Nixon years, led to a long, dark period where these drugs were lost to science, although they were still used in recreational and religious settings.

New studies have reignited the hope that psychedelics could be powerful medicines for mental disorders. In a *New England Journal of Medicine* report, two doses of the chemical from the psilocybin mushroom appeared as effective as six weeks of escitalopram (Lexapro), a standard antidepressant, for people with longstanding moderate-to-severe major depressive disorder (MDD). On many of the secondary measures of depression, such as remission of symptoms, psilocybin appeared better than the standard treatment, with 57 percent of subjects showing remission on psilocybin versus 28 percent in remission with escitalopram, although the authors caution that more research will be needed to confirm these results. In a study reported in *Nature Medicine,* MDMA was found to be more effective than placebo for people with severe posttraumatic stress disorder (PTSD). After three sessions with MDMA, 67 percent no longer met criteria for their diagnosis, and 33 percent showed complete remission, relative to 32 and 5 percent, respectively, after receiving a placebo.

Patients with disorders such as MDD and PTSD can certainly use more innovative, effective and safe treatments. Fewer than half of people with these disorders respond to medications or psychotherapy, and about a third of MDD patients have so-called treatment-refractory depression that fails to respond at all. Discovering new approaches to treat mental health conditions is critical. If psychedelics prove to be effective and safe for these disorders, they could be transformative in two interesting ways.

First, these medications appear to be effective after acute administration. Most psychiatric drugs, such as escitalopram for MDD, need to be taken for weeks, months or years to be effective. Presumably, they control symptoms but do not alter the disorder. When effective, psychedelics appear to confer long-term effects, sometimes

after a single administration, suggesting that they are not simply symptom-reducing but disease-modifying. Moreover, some results from previous studies suggest that the benefits are long-lived, although the duration of effects in these recent studies remains to be determined.

Second, the studies thus far have focused on psychedelic-assisted psychotherapy—not just the drug but the experience in the context of structured, time-limited psychotherapy. We know that psychotherapy can be helpful for most mental disorders, yet therapy is rarely combined with medication. Psychedelics remind us of the potential for combining medication and therapy, a practice that is infrequent when most antidepressants and antianxiety drugs are prescribed by primary care physicians. This combination is especially promising, given that some evidence suggests that psychedelics enhance neuroplasticity, thereby opening up a "critical window" of time to develop a healthier mindset.

These new studies with psilocybin and MDMA focus on only two of the many psychedelic compounds with potential for medical use. Could this class of drugs, which includes LSD, mescaline and many lesser-known or yet to be discovered chemicals, revive drug development for psychiatry? Perhaps, but we need to be mindful that in a previous era, overexuberance about psychedelics led to a backlash with little scientific research and no translation to medical use.

It has been 50 years since psychedelics were classified as Schedule I compounds: drugs with "high potential for abuse and no currently accepted medical use." New science suggests that we may need to revisit this classification. SA

Danielle Schlosser is a psychologist and senior vice president of clinical innovation at Compass Pathways, which is conducting clinical trials of psilocybin. In her prior role, she was in charge of vision and strategy for the behavioral health portfolio at Verily Life Sciences, an Alphabet company, and she remains on the faculty at the University of California, San Francisco.

Thomas R. Insel is a psychiatrist and neuroscientist who served as director of the National Institute of Mental Health from 2002 to 2015. He is an adviser to Compass Pathways, as well as to several digital mental health companies, and author of *Healing: Our Path from Mental Illness to Mental Health* (Penguin, 2022).

Does Consciousness Pervade the Universe?

A philosopher answers questions about "panpsychism"

By Gareth Cook

One of science's most challenging problems is a question that can be stated easily: Where does consciousness come from? In his most recent book, *Galileo's Error: Foundations for a New Science of Consciousness,* philosopher Philip Goff considers a radical perspective: What if consciousness is not something special that the brain does but instead is a quality inherent to all matter? It is a theory known as panpsychism. He answered questions from former longtime Mind Matters editor Gareth Cook.

For the complete interview, see www.ScientificAmerican.com/article/does-consciousness-pervade-the-universe/

Can you explain, in simple terms, what you mean by panpsychism?

In our standard view of things, consciousness exists only in the brains of highly evolved organisms, and hence it exists only in a tiny part of the universe and only in very recent history. According to panpsychism, consciousness pervades the universe and is a fundamental feature of it. This doesn't mean that literally everything is conscious. The basic commitment is that the fundamental constituents of reality—perhaps electrons and quarks—have incredibly simple forms of experience, and the very complex experience of the human or animal brain is somehow derived from the experience of the brain's most basic parts.

I should clarify that by "consciousness," I don't mean self-awareness or the capacity to reflect on one's own existence. I simply mean "experience": pleasure, pain, visual or auditory experience.

Human beings have a very rich and complex experience; horses less so, mice less so again. As we move to simpler forms of life, we find simpler forms of experience. Perhaps at some point the light switches off, and consciousness disappears. But it's at least coherent to suppose that this continuum of consciousness carries on into inorganic matter, with fundamental particles having unimaginably simple forms of experience.

What does panpsychism seek to bring to physics?

Philosophers of science have realized that physical science, for all its richness, is confined to telling us about the *behavior* of matter, what it does. Physics tells us, for example, that matter has mass and charge. These properties are completely defined in terms of behavior—things like attraction, repulsion, resistance to acceleration. Physics tells us absolutely nothing about what philosophers like to call the *intrinsic nature* of matter: what matter is in and of itself.

Consciousness, for the panpsychist, is the intrinsic nature of matter. There's nothing supernatural or spiritual, but matter can be described from two perspectives. Physical science describes matter from the outside in terms of its behavior. But matter from the "inside"—that is, in terms of its intrinsic nature—is constituted of forms of consciousness.

Do you foresee a scenario in which panpsychism can be tested?

You can't look inside an electron to see whether or not it is conscious, just as you can't look inside someone's head and see their feelings and experiences. We know that consciousness exists only because we are conscious.

Neuroscientists correlate certain kinds of brain activity with certain kinds of experience. We now know which kinds of brain activity are associated with feelings of hunger, pleasure, pain, and so on. This is really important information, but what we ultimately want from a science of consciousness is an explanation of those correlations. Why is a particular feeling correlated with a particular pattern of brain activity? As soon as you start to answer this question, you move beyond what can be, strictly speaking, tested, simply because consciousness is unobservable. We have to turn to philosophy.

Science gives us correlations between brain activity and experience. We then have to find the philosophical theory that best explains those correlations. In my view, the only theory that holds up to scrutiny is panpsychism. SA

Philip Goff specializes in the philosophy of mind and metaphysics at Durham University in England.

Getty Images

The most important stories about the universe and beyond

Journey from quantum to cosmic with *Scientific American Space & Physics*

sciam.com/space-physics

SCIENTIFIC AMERICAN
Space & Physics
ISSUE No.2
Here Come the Waves
GRAVITATIONAL-WAVE ASTRONOMY HAS MADE SOME STAGGERING DISCOVERIES—BUT EVEN MORE ARE ON THE WAY
WITH COVERAGE FROM nature
SCIENTIFIC AMERICAN
Space & Physics
ISSUE No.3
A Cosmic Crisis
AN EPIC DEBATE IS BREWING ABOUT HOW FAST THE UNIVERSE IS EXPANDING
Plus:
A NEW KIND OF NEUTRINO
Q&A WITH HAWKING'S FINAL CO-AUTHOR
GLITERING DIAMOND DUST IN SPACE
WITH COVERAGE FROM nature

Scientific American is a registered trademark of Springer Nature America, Inc.

Expertise. Insights. Illumination.

Discover world-changing science. Get 12 issues of *Scientific American* in print and explore our digital archive back to 1845, including articles by more than 150 Nobel Prize winners.

sciam.com/print&archive

Download on the App Store

GET IT ON Google Play

SPECIAL ISSUE : SEX AND GENDER
GENDER MYTHS DEBUNKED
UNHEALTH PRACTICE
SPEAKING IN WHISTLES
WHAT'S WRONG WITH COSMIC INFLATION?
SCIENTIFIC AMERICAN
WORKOUTS AND WEIGHT LOSS
Learn the surprising evolutionary reason why exercise alon won't shed pounds—and what to do about it
PLUS
PREVENTING THREATS TO CROPS
Airborne microbes' critical role PAGE 40
WHAT IS A KILOGRAM?
Science seeks new ways to ensure accuracy PAGE 46
MEDICAL SCANS
A shortage of rare atoms imperils disease imaging PAGE 68
SCIENTIFIC AMERICAN
AUTOMOBILE NUME
SCIENTIFIC AMERICAN
T. REX reexamined
SCIENTIFIC AMERICAN
September 1999

Scientific American is a registered trademark of Springer Nature America, Inc. Google Play and the Google Play logo are trademarks of Google LLC. Apple, the Apple logo, iPhone, and iPad are trademarks of Apple Inc., registered in the U.S. and other countries and regions. App Store is a service mark of Apple Inc.